Transcriptional Regulation of the Aggrecan Gene

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

By

Mr Ian Man Ho Li

September 2016

For the square peg in the round hole outsiders, your freedom is what they envy

Acknowledgments

After nearly three years here I have learnt that I know nothing in comparison to what there is left to understand, and that when you are happy, your work becomes brilliant it is never about how much you do, it is about joy from what you achieved and the fun. With that in mind I wish to thank my supervisor, Professor George Bou-Gharios, who although we have not always seen eye to eye, he has ensure that my thinking was unrestrained. He has committed an understanding to me of which without communication, collaboration and competition there is no requirement to work and I (now) understand that the environment we create fosters creativity. We've paid for this work with our time, time that is so easily lost and so deeply missed, I can say I am blessed to have worked on this project. I am grateful for Professor Peter Clegg, my second supervisor, who has provided thought provoking and invaluable discussions.

I am forever indebted to Dr Ke Liu, the silent wonderful person who has helped with all the transgenic experiments and carried out the oocyte injections. I am grateful for the support of the staff at the Biomedical services for ensuring the work ran smoothly. To Dr S. De Val, Dr A, Neal and Miss K. Fischer, I thank for teaching and helping me with the EMSA experiments. I wish to express gratitude to Dr B. Poulet and Dr L. Reynard for the conversations, ideas and encouragement during the final year of this work.

I am forever thankful for the support of my family, who, unbeknown to them allowed me to follow my heart and embark on an insane idea that I wasn't just part of a paint work and I could do something that I would enjoy (quite possibly for the rest of my life).

I believe the sine qua non is your emotions, don't let the footprints that remain after the dust settle define who you are, but the people who stood beside you. In light of this there are those whose paths have crossed and departed on their own adventure, who have made this experience richer and I am sure life has a funny way of ensuring we meet again.

"We are the conscientiousness of our own dreams, created at the point of time perfect for us; to collide, explode like fireworks and then vanish into mutual oblivion."

ABSTRACT

Aggrecan is a large- aggregating proteoglycan that is essential for the function of articular cartilage. A loss of aggrecan is one of the major event in osteoarthritis, a debilitating and degenerative joint disease. To understand how a gene is lost and what fails to prevent transcription in disease state, an understanding of the bases of transcription must first be sought. Transcription is the first step in gene regulation and the chromatin plays an important role in blocking or allowing transcription to occur. Enhancers are non-coding DNA sequences that allow the binding of proteins, such as transcription factors that drive transcription irrespective of distance or orientation.

Enhancers are marked by histone modifications that distinguish them from promoters and divide them into poised or active and are generally highly evolutionary conserved. Using publically available data on histone modifications found on ENCODE and transgenic mice this study has added to the understanding of the transcriptional regulation of aggrecan along with the known enhancer at -10kb. 3 intergenic regions, -35kb, -65kb and -87kb and one intronic (+26kb) from the transcription start site of aggrecan expresses primarily in the chondrocytes at E15.5. The -87kb enhancer marks chondrocytes that commit to an articular cartilage fate, and is strongly active in adult mice. The -65kb appears developmentally active, marking hypertrophic chondrocytes. The -35kb and +26kb are expressed in all chondrocytes at E15.5 mice and the -35kb is active in adult tissue. These enhancers bind to Sox9 and a loss of Sox9 in the -35kb enhancer shifts the expression from chondrocytes to fibroblast or perichondrium cells. These enhancers may play a role in response to diseases such as OA, as when these enhancers are transfected into differentiated murine ADTC5 cells and treated with Il-1β or hOSM there is a reduced level of expression from the -87kb element and an increase expression of the -65kb in response to Il-1β.

A highly conserved intronic sequences (+55kb) was identified that does not express in chondrocytes, rather in kidney and lungs and may play a role outside of aggrecan gene regulation.

Using the Acan promoter and the -35kb enhancer a cartilage-specific Cre recombinase line was generated providing a tool to explore the pathogenesis of cartilage genes in disease such as OA or in lineage tracing.

Table of Contents

Table of Figures

Index of Tables

ABBREVIATIONS

18S	18S ribosomal RNA	ENCODE	The Encyclopedia of DNA Elements
5mC	5-methylcytosine	EtBr	Ethidium Bromide
Ab	Antibody	EtOH	Ethanol
Acan	Aggrecan	FAIRE-seq	Formaldehyde-Assisted Isolation of Regulatory Elements followed by sequencing
ACAN	Aggrecan protein	FBS	Fetal bovine serum
Acan	Aggrecan Gene	FGF	Fibroblast growth factor
ADAMTS	a disintegran and metalloproteinase with thrombospondin motifs	FoxC1	Forkhead box C1
Bapx1	bagpipe homeobox homolog 1	g	gram
Bcan	Brevican	GAGs	Glycosaminoglycans
BMP	bone morphogenetic protein	*GAPHD*	Glyceraldehyde 3-phosphate dehydrogenase
bp	base pair	GEO	Gene Expression Omnibus
BSU	Biomedical Service unit	Gli2	GLI Family zinc finger 2
$CaCl_2$	Calcium chloride	GW	Gateway
Cas9	CRISPR associated protein 9	H3K27	Histone 3 lysine 27
CCD	Charge-couple device	H3K4	Histone 3 lysine 4
cDNA	complementary DNA	HA	hyaluronic acid
CFU	Colony forming unit	HIF-1α	Hypoxia-inducible factor 1-alpha
ChIA-PET	Chromatin Interaction Analysis by Paired-End Taq sequencing	HL	Hodgkin lymphoma
ChIP-seq	Chromatin immunoprecipitation with	HMM	Hidden Markov Model

	massively parallel DNA sequencing		
cm	Centimeter	HP1	heterochromatin protein 1
CmD	Cartilage matrix deficiency	*Hprt*	Hypoxanthine guanine phosphoribosyl transferase
Col10a1	Collagen Type X alpha I	HSP68	Heat shock protein 68
			High-temperature requirement
Col1a1	Collagen Type I alpha I	HtrA	A family of serine proteases
Col2a1	Collagen Type II alpha I	IGD	Interglobular domain
Col2a1	Collagen Type II alpha I gene	IGF	Insulin like growth factor
Cre	Causes recombination	Ihh	Indian hegdehog
CRISPR	Clustered regularly interspaced short palindromic repeats	Inr	Initiator
CRP	Complement regulatory protein- like	IP	Intraperitoneal injection
CS	Chondroitin sulphate	IVD	Intervertebral discs
			Journal of Visualized
CTCF	CCCTC-binding factor	JoVE	Experiments
DAB	3,3'-diaminobenzidine	K3Fe(CN)6	Potassium ferricyanide
DCE	downstream core element	K4Fe(CN)6	Potassium ferrocyanide
ddH2O	double distilled water	KCl	Potassium Chloride
dH2O	distilled water	KH_2PO_4	Potassium Phosphate monobasic
DNA	Deoxyribonucleic acid	KOAc	Potassium acetate
dNTP	Deoxynucleotide	KS	Keratin sulphate
DPE	downstream core promoter element	LARII	Luciferase Assay Reagent II
E.coli	Escherichia coli	LB	Luria-Bertani
ECM	Extracellular matrix	LCR	Locus control regions

EDTA	Ethylenediaminetetraacetic acid	LEC	lectin- like
EEO	low electroendosmosis	LICR	Ludwig Institute for Cancer Research
		LOD	Logarithm of the odds
EGF	epidermal growth factor like domain	*Matn1*	matrilin-1
eGFP	enhanced green fluorescent protein	Med1	Mediator
EGTA	ethylene glycol tetraacetic acid	MiRNA	micro ribonucleic acid
EMSA	Eletromobility shift assay	mL	millilitre

1. INTRODUCTION

1.1. Skeletal development

The formation and regulation of the axial and appendicular bones of the skeleton during adulthood can be considered as the end point of skeletal development. Evolutionary, it reflects the utilisation of different cellular sources and mechanisms to bring about the establishment of the structures that bring function. The formation is carried out in two processes: intramembranous and endochondral ossification. Mesenchymal progenitors have the potential to commit to the chondrocytic or osteoblastic linage depending on the restrictions applied to them and the developmental cues they receive (Tsang et al., 2015). All skeletal element form from one of three progenitors: the craniofacial derive from the cranial neural crest cells, limb and appendicular skeleton arise from lateral plate mesodermal cells and finally the paraxial mesoderm gives rise to the axial skeleton (Berendsen and Olsen, 2015). Bone formation follows defined steps: commitment to specialist cell linage, mesenchymal condensation, proliferation, extra-cellular matrix (ECM) disposition, mineralisation and maintenance.

1.1.1. Endochondral ossification

In embryonic development chondrogenesis occurs at the onset of skeletogenesis around embryonic day 9 (E9) in the mouse, and mesenchymal precursor cells commit to the chondrocyte phenotype by the signal cues from Paired box gene 1 (Pax1) and Pax9, which affects Nk3 homeobox 2 (Nkx3.2) and bagpipe homeobox homolog 1 (Bapx1) in the scleotome (Rodrigo et al., 2003). Sox9 a SRY-related high mobility group transcription factor is the most important factor during chondrogenesis, without its expression there is no commitment to the chondrocyte lineage (Bi et al., 1999). After initiation Sox9 expression is required for cell morphology and chondrocyte differentiation (Barna and Niswander, 2007). As the extracellluar matrix is deposited this is regulated by the sox trio: Sox9, Sox5 and Sox6. Pax1 is later downregulated as Runt-related transcription factor 2 (RUNX2) is expressed to induce hypertrophy of the chondrocytes at E13.5 in the mouse, Pax1 expressing progenitor mature to perichondrium cells and anlage of the intervertebral

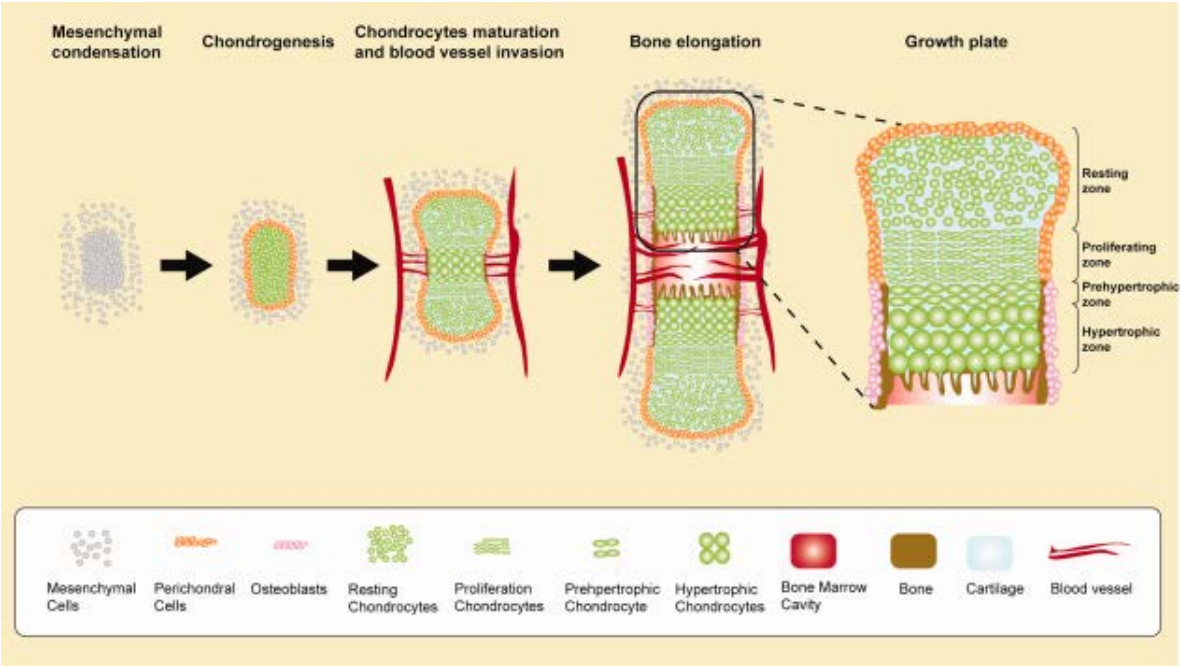
disks (Takimoto et al., 2013, DiPaola et al., 2005, Wallin et al., 1994). Scleraxis (Sxc) marks chondroprogenitors cells and aids Pax1 in the initiation of chondrogenesis, later Scx expressing progenitors form the bridge between cartilage and ligaments/tendons (Figure 2) (Sugimoto et al., 2013a, Sugimoto et al., 2013b, Cserjesi et al., 1995). Initially, the mesenchymal cells migrate to the site of future bone forming, the cartilage primordia, chondroprogenitors cells differentiate to chondrocytes to commence bone formation; they proliferate, mature and exit the cell cycle and lay down the outline of future bone. In the centre of the diaphysis of future long bones, the cells rapidly progress towards pre-hypertrophy. These then become hypertrophic and form an avascular cartilaginous scaffold that is gradually surrounded by a dense connective tissue known as the perichondrium (Goldring et al., 2006). The hypertrophic chondrocytes secrete matrix vesicles that promote calcification of the cartilage, marking the start of bone formation. Adjacent to the chondrocytes, pluripotent stem cells differentiate into osteoblasts; these are able to secrete the bone matrix that forms a bone collar that begins to surround the hypertrophic chondrocytes. The osteoblasts align in and form the start of the diaphysis of the bone. The primary ossification centre (POC) develops as blood vessels penetrate the periosteum and progress towards calcified cartilage. Eventually, the osteoblasts lay down the bone matrix that forms trabecular bone (Figure 1) (Goldring et al., 2006, Mackie et al., 2008).

Epiphyses form where the chondro-osseous junctions form that split the cartilage template. This leads to the formation of the growth plate, around E14.5 in the mouse defined by the four morphologically identifiable zones: resting consisting of round chondrocytes, that serves as a reserve of progenitor-like cells; the proliferating regions that has the characteristic longitudinal columns formation where round cells divide, flatten and arrange in the direction of longitudinal growth under the influence of the non-canonical Wingless (wnt) signalling pathway; the pre-hypertrophic is where the mature chondrocytes resides, the cells have left the cell cycle and undergo hypertrophic differentiation with the expression of indian hedgehog (*Ihh*) and collagen type X (*Col10a1*) (Yeung Tsang et al., 2014, Domowicz et al., 2009). The final zone, the hypertrophic, is divided into two sectors the upper and the lower, the chondrocytes expand in size in the

higher zone and in the lower zone cartilage calcifies to become bone (Goldring et al., 2006, Yeung Tsang et al., 2014). At this stage the fate of the chondrocyte is controversial, it has been suggested they undergo apoptosis or autophagy whereas other believe that they undergo differentiation and lineage switch into osteoblasts. However, by the use of a *Col10a1* cre recombinase lineage tracing experiment hypertrophic chondrocytes may switch lineages to osteocytes and osteoblasts and cre activated cells are detected in the POC and begin to express *Col1a1*, a marker for osteoblasts (Yang et al., 2014). From this point the growth plate expands and conventional bone growth, proliferation, matrix production and hypertrophy acts to increase the volume of the bone (Yeung Tsang et al., 2014).

Beneath the cartilage at the junction between the subchondral bone is separated by calcified cartilage, which persists after growth plate closure. Histologically is separated from the articular cartilage by the tidemark and is denser than the adjacent subchondral bone. It is uniquely adapted to transfer and distribute mechanical forces from physiological loading. Additionally, it helps to transform shear stresses into compressive and tensile stressing from the cortical plate during motion and joint loading (Milz and Putz, 1994, Burr and Gallant, 2012).

Figure 1: Endochondral bone formation



Bone formation follows distinct phases that are recognised by the cell morphology. Chondrocytes follow from precursors and eventually become pre-hypertrophic and hypertrophic as the blood vessels infiltrate to allow cartilage matrix degradation and osteoblasts to rise. The chondrocytes arrange in the direction of bone growth to form the growth plates. As this occurs osteoblast lay down and mineralises the bone matrix that surround the chondrocytes. The growth plate is made by four populations of cells that for the hypertrophic, prehypertrophic, proliferating and resting zone (Yeung Tsang et al., 2014).

1.1.2. Intramembranous ossification

Flat bones in the body such as those in the craniofacial skeleton (the calvaria, maxilla, and palate) are formed by intramembranous ossification. This occurs by the direct differentiation of mesenchymal cells into osteoblasts without the requirement of a cartilage template (Berendsen and Olsen, 2015). Neural crest-derived mesenchymal cells proliferate and aggregates into nodules, these cells commit to different linages, primarily osteoblasts but others will form vasculature such as capillaries. Osteogenic potential and

differentiation initiated in the mesenchymal cells via the signalling pathways fibroblast growth factor (FGF) and bone morphogenetic protein (BMP) and their target molecules Msh homeobox 1 (Msh1), FGF2 and Twist-related protein 1 (TWIST1) (Rice et al., 2000, Lenton et al., 2011). Vascular endothelial growth factor (VEGF) is essential in the development of the craniofacial skeleton and osteoblast-specific transcription factor osterix (Osx) are required for the commitment into intramembanous ossification (Duan et al., 2016, Duan et al., 2015). The osteoid forms between the forming blood vessels, the osteoid arise from the osteoblasts as they secrete collagen (mainly type 1) and proteoglycans that binds to calcium to allow mineralisation. As the osteoid becomes calcified, under the influence of TFs from the Wnt signalling such as Wnt16 (Jiang et al., 2014), osteoblasts move outwards leaving the matrix they secrete, dictating the direction of growth of the bone, some osteoblasts become trapped and become osteoclasts. Mesenchymal cells surround the osteoid and begin forming the periosteum creating layers of bone. Trabeculae in the periosteum thicken forming woven bone that is later replaced with lamellar bone and the vascular tissue become red bone marrow (Percival and Richtsmeier, 2013).

The two developmental processes of bone development create one functioning skeleton.Understanding the critical steps of these processes are aiding in the development of treatment for cartilage and bone repair, remodelling and diseases.

1.1.3. Postnatal skeletal development

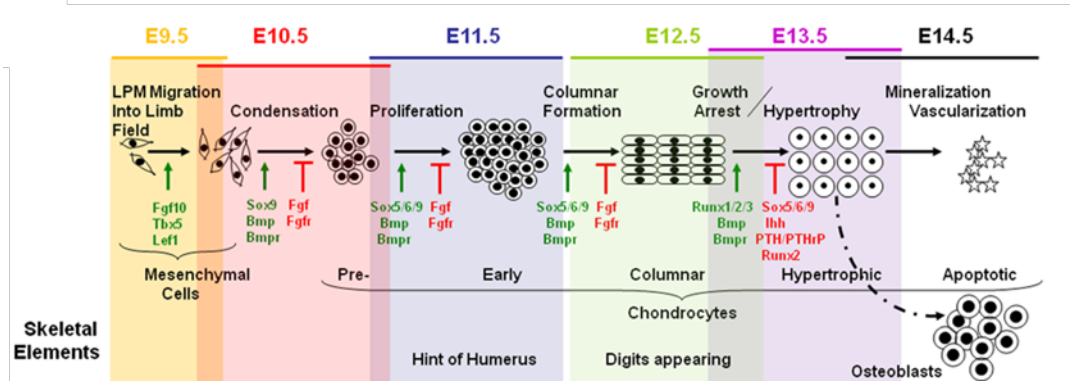
The epiphysis, metaphysis and diaphysis form the two ends and the shaft of the bone and the portion of these three sites increase with skeletal maturity. The longitudinal growth of the metaphysis and diaphysis continues after birth into childhood and adolescence. The epiphysis forms the spherical growth of the ends of the bone and shapes the articular surface that resides above it. It is also the site for the longitudinal growth by the expansion of the epiphyseal cartilage, resulting in defined regions.

Formation of the secondary ossification centre (SOC) appears in humans around 3months to 3 years of age. The generation of the SOC follows the same pattern as the POC, a

cartilage scaffold is formed and is mineralised, angiogenesis occurs and *de novo* synthesis of the bone matrix by osteoblasts and appears to be regulated closely by the same signalling pathways dictating embryonic endochondral ossification. However, during the formation of the SOC vascularisation of the cartilage occurs much earlier, before the matrix mineralisation compared to the POC (Alvarez et al., 2005, Blumer et al., 2005).

The growth plate chondrocytes reside on both ends of long bones and are essential for the continual linear growth of the bone through cartilage deposition following the direction of mechanical stimulus, until in human they close at puberty and the metaphysis fuses with the epiphysis (Dirckx et al., 2013, Nowlan et al., 2010, Albro et al., 2011). In mice growth plates do not disappear rather the growth slows down to an almost resting state. The growth plate is divided into distinct zones that undergo cell proliferation, ECM synthesis, hypertrophy and mineralisation following the direction of bone growth. The formation of bone and the resorption of bone by osteoclasts in the cortical endosteal and lateral metaphyseal surfaces are tightly balanced to allow the shaping and growth of the bone. The hypertrophic chondrocyte is important in fracture repair, as cartilage promotes the vascularisation and integration of bone into the damaged site following the same procedure as endochondral ossification, but with the possibility of failure (Bahney et al., 2014, Little et al., 2007). Additionally, the chondrocyte has the ability to transdifferentiate to osteoblasts during development, postnatally and during fracture healing (Zhou et al., 2014, Ono et al., 2014).

Figure 2: Limb development and the key transcription factors of each stage



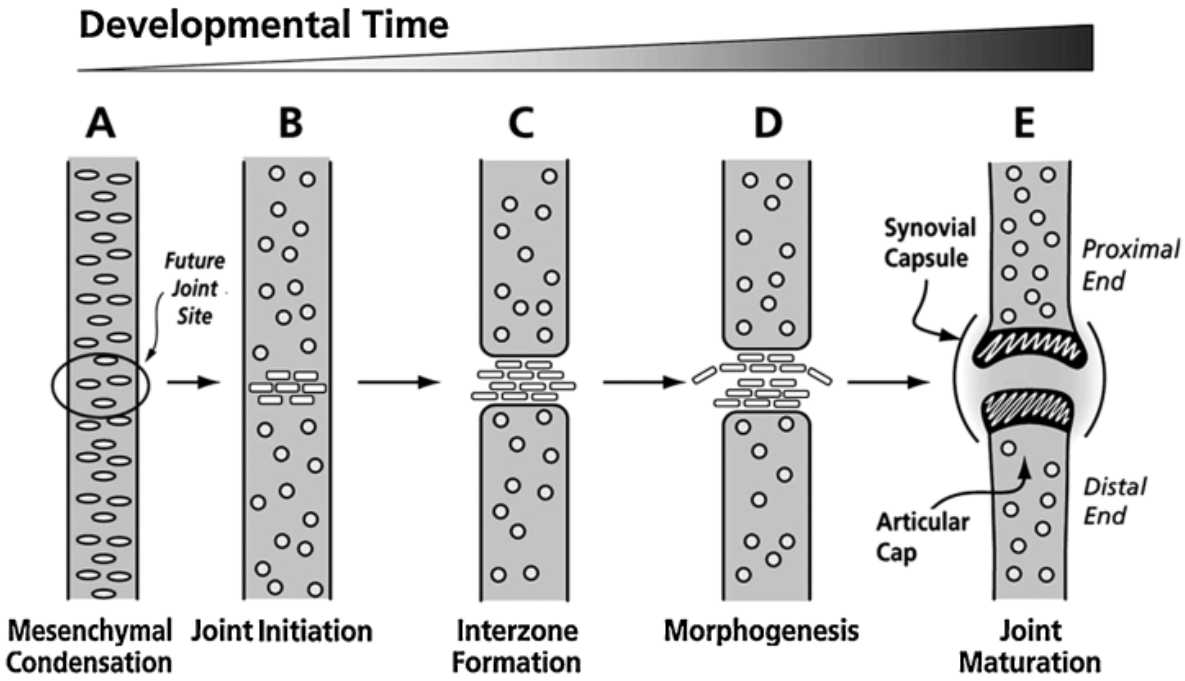
Mouse limb development with key events in the chondrocytes and the transcription factors that act on each stage to delay or allow progression (Taher et al., 2011).

1.1.4. Articular cartilage formation

Cartilage is an abundant connective tissue found throughout the body, the chondrocyte is the only cell-type found in this tissue and it is progressively replaced by bone during development. However, some may remain in adulthood to provide support and natural function of joints. The gliding surfaces of synovial joints are covered by articular cartilage, a highly specialised tissue. The plentiful ECM comprising of proteoglycans and collagen fibrils that provide tensile strength defines cartilage and its ability to resist compressive loads and allow friction free movement. The articular cartilage forms when mesenchymal cells that are undergoing differentiation into chondrocytes cease to express *col2a1* and an interzone starts and cells surrounding the forming joint express forming marked by growth and differentiation factor-5 (*Gdf5*) (Decker et al., 2015). In the mouse at E13.5 the dorsal and ventral cells of the developing joints express TGFβ type II receptor that remain in the localised zones to form the synovial lining, meniscal surface and groove of Ranvier (Figure 3) (Li et al., 2013). In the initial interzone cells that formed the intermediate compartment gives rise to the articular cartilage.

Postnatally, cells at the superficial region divide to allow for the articular cartilage to form and give raise to daughter cells that rapidly divide in the deeper zones that allows vertical growth along with the skeleton (Decker et al., 2014b, Decker et al., 2015).

Figure 3: Synovial joint formation



Mesenchymal condensations form (A) and as signals at the exact site of joint formation is initiated such as Hox cells gather and condense (B), the interzone forms (C) the interzone and epiphyseal cartilaginous tissue start moulding into the joint (D) and the components of the joint emerge with articular cartilage at each end (Pacifici et al., 2006).

1.1.5. Fate determination of chondrocytes

To regulate the lineage fate and bone development there are numerous signalling pathway ways that activate temporally. It is understood that the bone morphogenic proteins (BMPs) and transforming growth factor beta (TGF-β) signalling are required for the initiation of chondrogenesis. In mice at developmental day 10.5 (E10.5) the

mesenchymal cells in the limb buds aggregate to form condensed mesenchyme, the cell-cell association that occurs after cell aggregation and cluster formation are under the influence of BMPs and TGF (Barna and Niswander, 2007). Furthermore, BMP2 in a mouse embryonic fibroblasts culture induces chondrogenesis with a marked increase in the transcription factor (TF) High mobility group box member Sex-determining region Y- box 9 (Sox9) expression. In mice, at E10.5, BMPs inhibit chondrogenesis whereas at E11.5 there is transient activation of the TGF-β signalling as chondrogenesis proceeds. Smad4 is common in both pathways and is essential in limb skeletogenesis (Benazet et al., 2012, Yan et al., 2014, Lorda-Diez et al., 2013, Retting et al., 2009).

Sox9 along with L-Sox5 and Sox6 are a trio of TFs that are required and sufficient for chondrogenesis (Zhou et al., 1998, Ikeda et al., 2004, Bi et al., 1999). The Soxs have a Sry-related HMG box domain which both binds and bends DNA. The protein contains additional sites including the dimerisation domain that allows binding of multiple sites on the DNA. The gene is switched on in mesenchymal cells expressed in prechondrocytes and switched off during the prehypertropyic stage of chondrogensis (Bi et al., 1999). The Sox trio orchestrates the organisation of the whole transcriptional machinery formation at the promoter. Sox9 is essential for chondrogenesis and cell determination of the chondrocyte lineage; indeed it is coined as the master regulator of chondrogensis (Bi et al., 1999, Bi et al., 2001, Dy et al., 2012). In the condensed mesenchyme Sox9 maintains cell morphology and promotes chondrocyte differentiation (Barna and Niswander, 2007). To show its importance, heterozygous human mutations in SOX9 results campomelic dysplasia a severe skeletal disorder characterised by poorly formed tracheal and rib cartilage (Akiyama et al., 2002, Akiyama et al., 2004, Bi et al., 2001). RUNX2 is essential for osteoblast lineage commitment. It is expressed co-ordinately with SOX9 and β-catenin (Akiyama et al., 2004, Clevers, 2006). WNTs/ β-catenin suppress Sox9 mediated differentiation and promote RUNX2 oestrogenic differentiation. The Hox proteins are essential in the developing embryo they are characterised by their ability to bind to the homeobox motif of DNA (Minguillon et al., 2012). The Hox-A and Hox-D are essential for limb patterning whereas HoxD 11-13 and HoxA13 regulate RUNX2 expression, therefore of importance in skeletal development (Kmita et al., 2005, Minguillon et al., 2012). BMPs,

WNTs and IHH are generated in the perichondrium cooperate or antagonise cell growth and the ability of cells to leave the cell cycle (Koyama et al., 2007). The non-canonical BMP signalling pathway plays a part in the maintenance of the growth plate (Retting et al., 2009). Additionally, the transcription factors human short stature homeobox-containing gene (SHOX) has been identified to regulate limb bud development especially in the growth plate formation and the transition into mature chondrocytes (Beiser et al., 2014, Bobick and Cobb, 2012, Yu et al., 2007). IHH is needed for differentiation of round proliferative chondrocytes from the resting zone to the column forming chondrocytes. Additionally, the Gli transcription factors (1, 2 and 3) are important in mediating IHH signalling (Gualeni et al., 2010, Koziel et al., 2005, Hilton et al., 2005). They regulate proliferation and hypertrophy via a negative feedback loop with the parathyroid hormone-related protein (PTHrP) signalling (Mak et al., 2008, Hilton et al., 2007). This negative feedback determines the length of the proliferating zone and prevents premature hypertrophy, with PTHrP delaying the initiation of hypertrophic differentiation consequently regulating the size and length of the growth plate. Additionally, at this stage is it critical that the TF forkhead box C1 (FoxC1) direct and indirect activities, especially its interaction with Gli2 guide *Ihh* pathway to drive hypertrophy, as a specific loss of FoxC1 in chondrocytes results in abnormal skeletal development and death shortly after birth in knock-out mice (Yoshida et al., 2015). In the proliferating zone chondrocytes divide in a column formation in the direction of bone growth, this spatial organisation is critical for the formation of normal bones. This is under the regulation of the non- canonical WNT signalling pathway allowing for planar cell polarity and organisation. Hypoxia inducible factor 1α (HIF-1α) has been shown to be active in chondrocytes (Lafont et al., 2007). It is required for ECM synthesis, chondrocyte survival and differentiation (Lafont et al., 2008, Lafont et al., 2007, Zhang et al., 2011). Notch signalling (Hilton et al., 2008).

The final stage is the commitment of hypertrophic differentiation, on leaving the cell cycle they enter pre-hypertrophic state and initiate the hypertrophic increase in size and cell shape. In the second stage of hypertrophy there is a swelling with no increase in mass. This stage may be partly dependant on IGF signalling (Cooper et al., 2013).

1.1.6. The chondrocyte environment: the ECM

The chondrocyte in the articular cartilage secrete ECM that sustains the cartilage this plays an important structural and signalling medium for the cells, each chondrocyte is surrounded by a pericellular matrix (Akkiraju and Nohe, 2015). The ECM is highly ordered from the superficial layer to the deeper zones of the articular cartilage that can be divided into four zones: superficial, transitional, deep and calcified. The superficial contains collagen fibrils, collagen type II IX and XI and low levels of proteoglycans with the chondrocytes are elongated in shaped aligned parallel to the surface. This resists shearing, compressive force and tensile forces applied by locomotion (Yu and Urban, 2010, Gao et al., 2014).

The transitional zone contains more proteoglycan therefore withstands compression better, it contains less collagen than the superficial that are organised in bundles or layers. The chondrocytes here are round and make up the majority of the height of the transitional zone. In the deep zones the chondrocytes share similar morphology to the transitional zone but aligned perpendicular to the articular surface with the largest composition of collagen that extend to the bone with proteoglycan content, especially aggrecan, is the highest here. The calcified zone is divided by the tidemark and acts as mechanical and physical interface between to the hyaline cartilage and the subchondral bone. Chondrocytes here are scarce and hypertrophic, the main function of this zone is to act as an anchoring point for the collagen fibrils (Cromar et al., 2014).

The ECM is made up of collagens, proteoglycans, hyaluronan, link proteins and other smaller molecules. Collagens are the most abundant macromolecules that stabilises the matrix with the main type being type II. Proteoglycans make up the second largest group of macromolecules such as aggrecan, decorin and biglycan. (Akkiraju and Nohe, 2015, Belluoccio et al., 2006). The composition is the ECM is outlined in Table 1.

Table 1: The composition of the extracelluar matrix

Protein	**Function**
Collagen	
Type II	Tensile strength
Type IX	Tensile properties and interfibrillar connections
Type XI	Fibril formation
Type VI	Microfibrillar network by binding to hyaluronan, biglycan and decorin
Type X	Endochondral ossification
Type XII	Associated with collagen fibrils
Type XIV	Associated with collagen fibrils
Proteoglycans	
Aggrecan	Compressive stiffness
Versican	Calcium-binding and selectin-like properties
Perlecan	Cell-matrix adhesion
Biglycan	Binds collagen VI and TGF-β
Fibromodulin	Fibrillogenesis and interacts with type II collagen fibrils
Decorin	Binds collagen II and controls size and shape of collagen fibrils
Lubricin	Dose-dependent boundary lubricant of cartilage
Chondromodulin-1	Glycosylated transmembrane protein. Inhibits angiogenesis

Proline arginine-rich end leucine-rich repeat protein (PRELP)	Anchors basement membranes to connective tissue
Osteoadherin (OSAD)	Restricted to mineralised tissue, possible signalling molecule
Chondroadherin (CHAD)	Cell to matrix interaction by α2β1 integrin and cell surface proteoglycans
Others	
COMP	Stabilises collagen networks and fibril assembly
Hyaluronic acid	Interacts with aggrecan to allow retention
Link protein	Stabilises aggrecan G1 domain for hyaluronic acid
Fibronectin	Cell adhesion
Laminins	Glycoproteins, cell adhesion
Matrilin-1	Ssh signalling, mechanical response, formation of filamentous network
Matrilin-3	Binds to Collagen IX to form filamentous network
Tenascin-C	Extracellular matrix glycoprotein, modulates cell adhesion, integrin ligand

The components of the ECM and cartilage (Belluoccio et al., 2006, Wilson et al., 2010, Sophia Fox et al., 2009).

1.2. Aggrecan

1.2.1. The Lectican protein family

Lecticans (also known as hyalectans) are family of proteoglycans that are found in the ECM and are only found in animals of a vertebrate lineage. There are four members of the family: Aggrecan (ACAN), Brevican (BCAN), Neurocan (NCAN), and Versican (VCAN), they are characterised by sharing similar C- and N- globular domains separated by a glycosaminoglycan attachment region (Kresse and Schonherr, 2001, Shibata et al., 2003).

Each lectican has a characteristic distribution pattern in tissues. VCAN is ubiquitously expressed but mainly found in connective tissue, ACAN is the most abundant in cartilage, BCAN and NCAN are nervous system-specific.

VCAN has hygroscopic properties that allow to help create a loose and hydrated matrix, it has a wide tissue distribution found in loose connective tissue fibrous, articular and elastic cartilage as well as in neuronal tissue. VCAN has roles in cell adhesion, proliferation, migration and ECM assembly. BCAN is one of the most important neural proteoglycan, unlike ACAN is does not contain a G2 domain and remains in a non-proteoglycan state. BCAN localised to the outer surface of neurons at perisynaptic sites and aids in the formation of the perineuronal net (Frischknecht and Seidenbecher, 2012). NCAN is expressed in the subplate and marginal zone of the brain and may play a role in neuronal migration and formation of axonal fibres (Maeda, 2015).

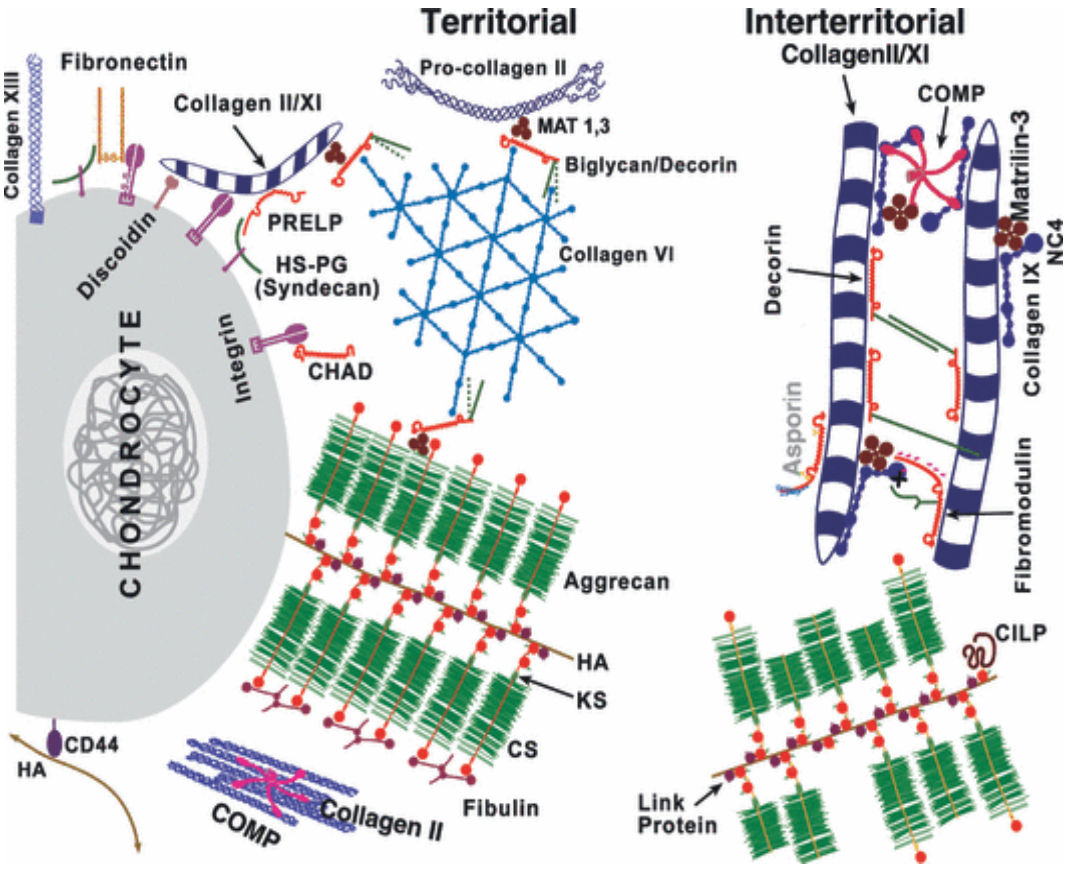
The link proteins, also known as Hyaluronan and proteoglycan link proteins (HAPLN), are within close genomic proximity to the Lectican family genes. The link family stabilises the hyaluronic acid binding to the G1 domains of the Lectican family proteogylcans to allow aggregates to form. However, the close proximity of the link proteins to the proteoglycan in the genomic context does not mean they function together. For example HPALN3 is widely expressed in the heart, mammary gland, ovary, lymph node, spleen, thymus and lungs but ACAN is restricted to the aorta and trachea. Furthermore HAPLN1 is not co-expressed in the same regions as VCAN (Spicer et al., 2003). HAPNL1 interacts with both

ACAN and VCAN but with the latter more strongly (Shi et al., 2004, Matsumoto et al., 2003).

1.2.2. The aggrecan protein

The large aggregating chondroitin sulphate proteoglycan, aggrecan (ACAN), is the main structural proteoglycan found in the ECM, the main isoform of ACAN is around 240-270kD (Vilim and Fosang, 1993). It dictates the function of cartilage by the large number of glycosaminoglycans (GAGs) that surrounds it chondroitin sulphate (CS) and keratin sulphate domains (KS) (Valhmu et al., 1995, Kiani et al., 2002). HA is a large glycosaminoglycan that acts as a long-chain biopolymer in solution. HA facilitates in the retention of ACAN by forming aggregates stabilised by link proteins. The ACAN-HA complex is retained at the cell surface by connections with HA with HA synthase or HA bound to CD44. The GAGs are highly negatively charged that attracts positively charged sodium ion from the tissue fluid which in turn retains water molecules into the local area due to a large osmotic swelling pressure that is created and extremely high fixed charge density (Kiani et al., 2002, Roughley, 2006, Johnson et al., 2014). ACAN is too large to redistribute and therefore forms a compression resistant gel. Additionally, ACAN swells in an aqueous environment limited by the collagenous framework which is in equilibrium and provides the tensile strength of the system. When a compressive force is applied water is displaced and ACAN concentration increases, once this load is removed the equilibrium is restored (Watanabe et al., 1998, Kiani et al., 2002). ACAN expression is well conserved in evolution and has been described in xenopus, mammals and other animals (Doege et al., 1991, Walcz et al., 1994, Li and Schwartz, 1995, Doege et al., 2002).

Figure 4: The chondrocyte environment



The chondrocyte is the only cell type found in cartilage it produces all the nesscary elements of the matrix that gives cartilage its function. Aggrecan is found both in the territorial and interterritorial regions, it is the most abundant protein. It is reliant on the GAGs (green) that attaches to its CS and KS domains to keep the osmotic balance in cartilage. Collagen forms a network surrounding the chondrocyte that prevents the movement of aggrecan and the tension from the fibrils balances the osmotic pression created by the GAGs (Heinegard, 2009).

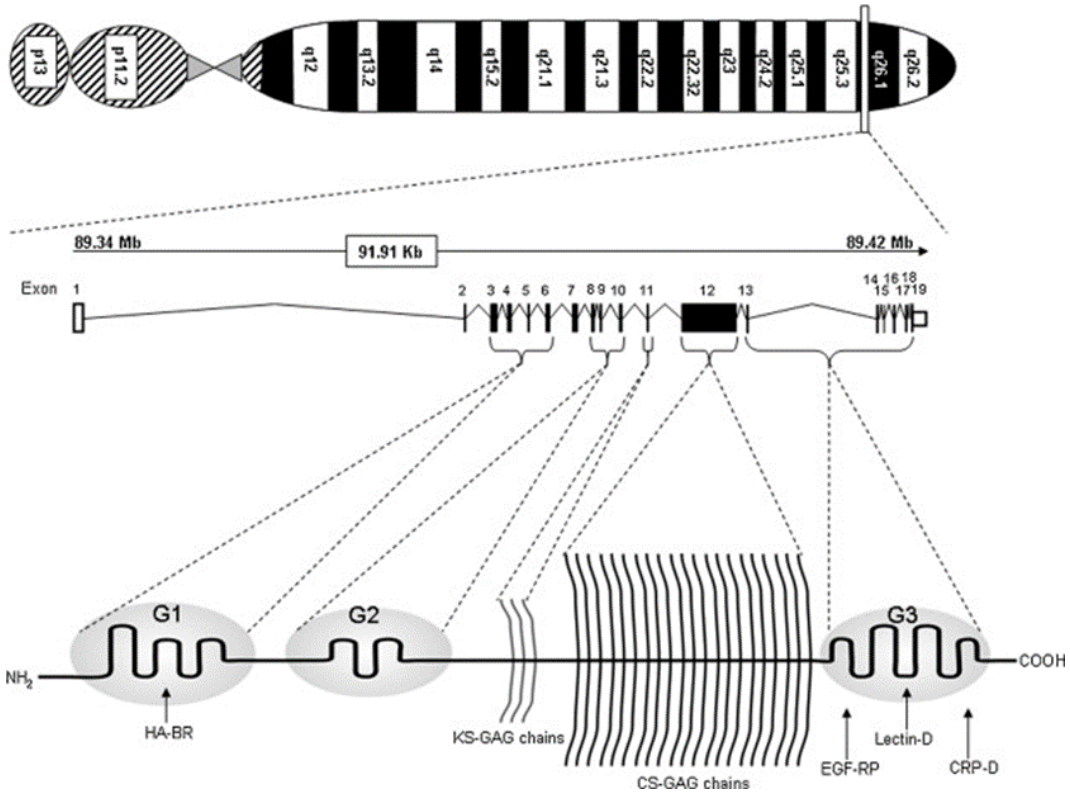
1.2.3. The ACAN protein structure

ACAN is synthesised as a 370kDa core protein precursor that undergoes post-translational modifications to make the mature 270kDa proteins and other isoforms. The protein structure of ACAN has three globular domains, one interglobular domain (IGD) and two GAGs attachment domains; CS and KS located between G2 and G3. The two N-terminal

globular domains, G1 and G2, show disulphide-bonded structural motifs (Figure 5). G1 consist of three loop-like subdomains- loop A, B and B' which are similar to link protein. The B and B' loops form a tandem homologous repeat that is critical for hyaluronic acid binding activity and is therefore important in the formation of the ECM network. G2 lacks HA binding activity. Both G1 and G2 are highly conserved. The IGD is susceptible to cleavage by proteinases including matrix metalloproteinases (MMPs) and aggrecanases. The G2 domain is composed of disulphide-bonded B and B' domains but with the inability to interact HA. KS domain allows GAG attachment composed of 6-residue amino acid repeats of proline and serine. The can be replaced by O-linked oligosaccharides or KS chains. The CS domain follows the KS domain which is divided into two CS1 and CS2 differentiated by their amino acid composition. CS1 has 19 amino acid repeats of two glycine-serine sequences, in humans this varies from 13-33 repeats; the reason for this is unclear. CS binds to the serine residues on GAGs. The CS2 has a variable amount of amino acids that allows the random distribution of CS chains. However, this leaves the region susceptible to cleavage from aggrecanases. The CS domain is the largest domain and is locate in the middle of the aggrecan core protein. It is divided into two subdomains CS1 and CS2; it is coded for on a single exon, the CS chains consist of repeating disaccharides units of *N*-acetylgalactosamine and glucuronic acid, the *N*- acetylgalactosamine carbon 6 and 4 are the sites for sulfation. Finally, the G3 domain is essential for the intra-cellular trafficking and secretion of the molecule. It is composite of three structural motifs: EGF-like, Lectin-like and CRP-like. Its lectin domain is able to interact with other ECM proteins *in vitro*. The lectin-like domain has been found to activate the complement pathway (Melin Furst et al., 2013). Due to proteolysis, the G3 domain is normally cleaved in the mature aggrecan molecule. Hyaluronan is a ubiquitous disaccharide chains and are non-covalently bound to ACAN at the B and B' domains to form aggregates (Watanabe et al., 1997). After synthesis ACAN is secreted as monomers into the ECM and binds to hyaluronan by the intermediate link protein and 1 covalent link with matrilin-1 (Heinegard, 2009, Cao et al., 1998), this allows the chondrocyte's ECM to form distinct regions the territorial and interterritorial where the collagen fibrils form a mesh that

ACAN is too large to pass through localising the protein to the chondrocyte to confer its function (Figure 4) (Heinegard, 2009, Goldring, 2012).

Figure 5: Human aggrecan protein and gene organisation



The human aggrecan gene composes of 19 exons that code for the mature protein. The globular domains allows for the large size of the protein whereas the subdomains such as the KS and CS are sites where GAG attach. These sites are functionally important in preserving the ability of aggrecan to draw water into the cartilage (Morawski et al., 2012).

1.2.4. Gene organisation of ACAN

ACAN has now been fully sequenced and annotated in multiple genomes, for instance the human (Valhmu et al., 1995, Doege et al., 1991), mouse (Walcz et al., 1994, Watanabe et al., 1995, Krueger et al., 1999), rat (Doege et al., 1994), bovine (Hering et al., 1997) and more recently the equine (Caporali et al., 2015), throughout these species the gene

organisation remains conserved with the exception of the splice sites and variants derived from the G3 domain coding regions and the length of the introns (Table 2). The mouse aggrecan gene (*Acan*) spans more than 61kb and contains 18 exons, whereas the human aggrecan gene has been mapped to 15q26 containing 19 exons and 18 introns. The exon-intron organisation of human and rat *Acan* are similar compared to the murine counterpart (Valhmu et al., 1995). In *Acan* for both mouse and human, exon 1 codes for the 5' UTR and translation begins at exon 2, intron 1 is the largest intron being 21kb. Analysis of gene structure revealed that the B loops of G1 and G1 domains are encoded by two exons: B of G1 exons 4 and 5, B of G2 exon 8 and 9. B' is encoded on 6 (G1) and exon 10 (G2).

Table 2: Gene organisation and comparison between human and mouse *Acan*

	Human				Mouse			
Domain	Exons (bp)		Introns (kb)		Exons (bp)		Introns (kb)	
5' UTR	1	375	1	13.0	1	509	1	21.0
Signal peptide	2	77	2	2.4	2	77	2	8.5
G1-A	3	384	3	1.0	3	384	3	1.8
G1-B	4	175	4	1.6	4	175	4	2.0
	5	128	5	1.5	5	128	5	1.6
G1-B'	6	294	6	1.9	6	294	6	1.4
IGD	7	378	7	1.5	7	405	7	1.4
G2-B	8	175	8	0.5	8	174	8	0.19
	9	128	9	1.4	9	129	9	1.5
G2-B'	10	294	10	2.1	10	294	10	3.0
KS1	11	240	11	2.9	11	216	11	1.2
KS2/CS	12	240	12	0.9	12	3482	12	4.7
EGF1	13	114	13	8.4	13	114	13	2.8
EGF2	14	114	14	3.2	-	-	14	0.55

LEC	15	159	15	0.5	14	158	15	0.7
	16	83	16	0.8	15	83	16	0.9
	17	145	17	0.8	16	145	17	0.6
CRP	18	183	18	0.5	17	183	-	
3'UTR	19	212	-	-	18	74	-	

Comparison of gene organisation of aggrecan between human and mouse. The intron length between species varies although the coding sequences give similar coded proteins. The EGF2 within the G3 of the mouse aggrecan is not coded for. G1, G2 and G3 are the globular domains, keratan sulphate (KS) chondroitin sulphate (CS), and epidermal growth factor like domain (EGF), lectin- like (LEC) and complement regulatory protein- like (CRP) (Krueger et al., 1999, Watanabe et al., 1995, Valhmu et al., 1995, Walcz et al., 1994).

IGD is located between the two globular domains and is coded for on solely exon 7 and is susceptible to proteinases. KS domain is located on the C-terminus of G2 and is coded for by exon 11 in mice and by exon 11 and the 5' end of exon 12 in humans. Exon 12 encodes the CS domain. G3 region consist of two alternatively spliced EGF-like domains, a LEC-domain and a CRP-like domain. Exon 13 and 14 encodes EGF-like, Exon 15, 16 and 17 for LEC-like region, only the exon 16 is capable of secretion of ACAN product (Luo et al., 1996)and CRP-like by exon 18. In the mouse the EFG-like is coded for by exon 13 (Valhmu et al., 1995). The LEC-domain is coded for on exons 14, 15 and 16, and the CRP-like is encoded on exon 17 and 18 (Watanabe et al., 1995). Finally the extra exon in the human *Acan*, exon 19 codes for the 25 amino acid C-terminus of aggrecan core protein and the 3' URT (Valhmu et al., 1995).

1.2.5. Gene expression of Aggrecan

ACAN has been found in various cartilage types such as in the growth plate and articular cartilage (Domowicz et al., 2009) as well as in developing limbs and bones, being expressed late in development. Additionally, ACAN is essential for the function of the intervertebral discs (IVD), located in the nucleus pulposus and it is contained by the collagen fibrils of the annulus fibrosis (Colombier et al., 2014, Melrose et al., 2008, Neidlinger-Wilke et al., 2006).

Additionally, it has been suggested that ACAN is involved in cardiogenesis. In chicks, at day 7 of development the aortic wall is composed of the endothelium with little or no ACAN staining. By day 14 intimal thickening was clear a weak ACAN immunolabeling was observed. This was localised in the lamellar and interlamellar cell layers. At day 21 ACAN was still limited to the outer regions. Immunolabeling showed that ACAN had a restricted distribution across the aortic wall in chickens and ACAN content is low at all stages (Arciniegas et al., 2004, Zanin et al., 1999). It is further found in the endocardial cushion tissue cell, epicardial cells of the heart as well as the carotid artery and aorta (Lincoln et al., 2006b, Lincoln et al., 2006a, Fang et al., 2014).

There are other ACAN isoforms, such as those found lacking the GAG side chains, reside in the central nervous system found in the adult cerebral cortex, deep cerebellar, the colliculi, the spinal cord, dentate gyrus and in the hippocampus in rats (Domowicz et al., 2008, Yamaguchi, 2000, Viapiano and Matthews, 2006). Other areas where ACAN has been identified include during liver fibrogenesis (Krull and Gressner, 1992) and in adipose (Voros et al., 2006).

1.2.6. Aggrecan gene mutations

Due to the importance of ACAN it is not surprising that *Acan* mutation associated diseases are uncommon due to lethality. However, when they do occur they result in severe skeletal defects, chondrodysplasias and early onset osteoarthritis (OA) (Gleghorn et al., 2005, Aspberg, 2012). To highlight the impact of genetic mutations in *Acan,* a missense mutation formed from a switch from G to A in exon 17 coding for the LEC-domain from the G3 domain results in a change in amino acid from valine to methionine (V2303M), this disrupts the interactions in the ECM and has been suggested to disorganise the ECM to cause the rare disease, autosomal dominant familial osteochondritis dissecans (Stattin et al., 2010). A disease characterised by the disassociation of the cartilage and subcondral bone from the surrounding tissue, disturbed chondroskeletal development, early OA and osteochondritic lesions in joints but doesn't affect the translation of ACAN as normal levels of protein is detected compared to wildtype (Stattin et al., 2010). Another mutation in the LEC domain of the G3 domain results in an autosomal recessive form of

spondyloepimetaphyseal dysplasia, a disease characterised by severe short statue, brachydactyly. Radiographical examination reveals irregular epiphyses with widened metaphyses, spinal abnormalities such as platyspondyly and cervical-vetrabral clefts. The point mutation causes a substation of aspartic acid with asparagine (D2267N) obliterates one of the two calcium coordinates in the LEC-domain and creates an N- linked glycosylation site but is less disruptive then the V2303M mutant (Tompson et al., 2009). Whereas a failure in sulfation of ACAN can result in chondrodysplasia such as mutations in the enzyme Impad1, causes a loss of function leading to impaired sulfation of chondroitin on ACAN (Vissers et al., 2011). Histologically, in the mouse mutations in *Impad1* shows delayed and disorganised maturation of the growth plate chondrocytes and no expression of ACAN in the joints (Sohaskey et al., 2008). Three loss of function heterozygous mutations in *Acan* results in advance bone age, early OA and short stature (Nilsson et al., 2014). Mutations c.2026+1G>A, c706T>C missense mutation in exon 14 causes a protein switch of L2355P, and a frame shift c.272delA have different outcomes to ACAN. The first interrupts a highly conserved splice donor site leading to a disrupted protein sequence, the second in the LEC was predicted to perturb protein function and the third results in a truncated protein (Nilsson et al., 2014). This result in haploinsufficiency with the exception of the second with varying severity of disease. Lastly, a failure of CS attachment or defective CS chain length results in short stature, delay cartilage formation and severe chondrodysplasia (Sato et al., 2011, Watanabe et al., 2010, Hiraoka et al., 2007, Schwartz and Domowicz, 2002). For example, a single nucleotide deletion or compound heterozygous nonsense mutation in the SLC35 gene family that codes for nucleotide-sugar transporter SLC35D1 in humans results in Schneckenbecken dysplasia (Hiraoka et al., 2007). SLC35D1 is involved in CS biosynthesis and a loss reduces the chain length of CS and amount of chains on ACAN. Schneckenbecken dysplasia manifests itself with severe skeletal defects, a distinctive snail-like of the ilia from medial bone projection from the inner iliac margin, thoracic hypoplasia, flattening of the vertebral bodies and short, thick bones (Hiraoka et al., 2007). *Acan* mutations do not all affect the cartilage, Hodgkin lymphoma (HL) is a lymphoid neoplasm from B-cells, it is the most common cancer in adolescents with it aetiology poorly understood. A family with three of the five children

was diagnosed with classical HL and their blood was sequenced for a possibly genetic causes (Ristolainen et al., 2015). A homozygous deletion of *Acan* was found in all affected individuals and both parents and one unaffected child were heterozygous for the c2836_2892del which results in a 19 amino acid deletion (Ristolainen et al., 2015). It is thought that the microenvironment of the cancer cells promotes oncogenic potential, however due to heterogeneity of cancer, this deletion was part of a panel of 10 possible causative genes. Other interesting disorders arise from abnormalities in ACAN include schizophrenia (SZ) and biopolar disorder (BP), in the brain ACAN is localised to the perineuronal nets (PNN) and glial cells but in SZ and BP there was a decrease in ACAN in PNN and glial cells in (Pantazopoulos et al., 2015). Although not all variation in *Acan* are detrimental but beneficial in selective criteria and keeping genetic variation open in the population, for example a study conducted by Weedon *et al* in 2008 identified the SNP rs8041863 which appears to be one of the 20 important loci to determine the height of an individual (Weedon et al., 2008).

In mice, mouse cartilage matrix deficiency (cmd) is naturally occurring disorder. It is caused by a 7bp deletion in exon 5 of *Acan* (Krueger et al., 1999). Autosomal recessive mutation characterised by cleft palate, short limbs, tail and snout. Heterozygous mice present normally whereas homozygous mice die of respiratory failure. This arise from poor formation of the tracheal cartilage resulting in the collapsed trachea (Watanabe et al., 1994).

The *Acan* point mutations model the nanomelic chick is a lethal autosomal recessive genetic lesion, present with reduced trunk, head size and skeletal malformation due to reduced ACAN in the ECM (Domowicz et al., 2009, McKeown-Longo and Goetinck, 1982, Yamagata et al., 1993, Domowicz et al., 1996, Stattin et al., 2010, Li et al., 1993, Vertel et al., 1994). Nanomelic chondrocytes produce a truncated core ACAN protein precursor of 300kDa, but fails to process the precursor at the endoplasmic reticulum and translocate the product (Vertel et al., 1993).

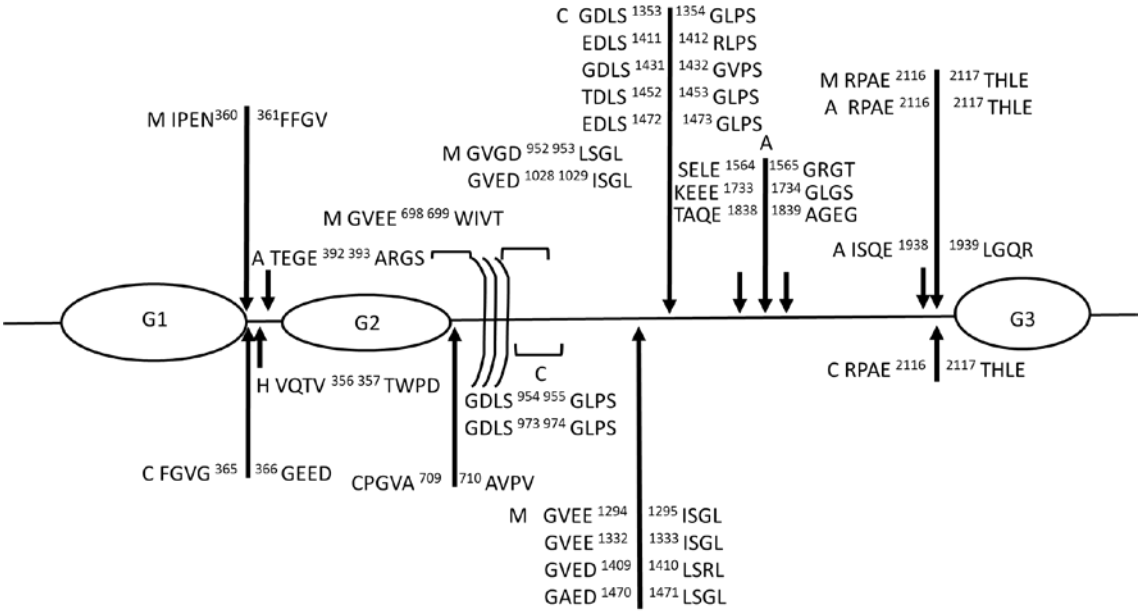
In the Dexter and Scottish highland cattle Bulldog dwarfism is single locus disorder derived from a 4bp insertion in exon 11 (2266_2267insGCCA) known as $ACAN^{BD1}$ and a rarer -

198C>T transition in exon 1 ($ACAN^{BD2}$) and follows classical Mendelian inheritance (Catalina Cabrera et al., 2016, Cavanagh et al., 2007). Homozygous present with severe chondrodysplasia of disproportionate dwarfism from abnormal cartilage formation, short vertebral column, micromelia, a large head with retruded muzzle, cleft palate, protruding tongue and die within the 7^{th} month of gestation. Heterozygous cattle have a milder form of dwarfism. $ACAN^{BD1}$ and $ACAN^{BD2}$ results in nonsense mediated decay even heterozygous cattle have 8% mRNA levels of *Acan* when compared to their wildtype counterpart due to both variants of the disease producing premature stop codon in the mRNA (Cavanagh et al., 2007).

1.2.7. Proteolysis of Aggrecan

ACAN is sensitive to proteolysis at sites in the IGD between the N-terminal G1 and G2 globular domains (Figure 6). MMPs, found in cartilage and bone have been shown to have the ability to degrading proteoglycans at neutral pH. MMP-3 was revealed in human articular cartilage to cleave at the Asn^{341}- Phe^{342} bond in the IGD. Other MMPs -1, -2, -7 -8, -9 and -13 have been linked to the proteolysis of ACAN at the same site or sites closer to the C-terminus (Troeberg et al., 2014, Troeberg and Nagase, 2012). In OA patients the majority of ACAN fragments are not due to MMP activity but at the Glu^{373}-Ala^{374} bond in the IGD (Sandy et al., 1992). This site is the work of aggrecanases, the first purified A Disintegrin And Metalloproteinase with Thrombospondin motif 4 (ADAMTS-4), then aggrecanase 2 or ADAMTS-5. The ADAMTS family of the metzincin superfamily of metalloendopeptidases have broad catalytic activities, they are proteinases that are both secreted and have extracellular potential. They have a common domain organisation of N-terminus, signal peptide, pro-domain, catalytic, disintegrin-like domain, central therombospondin repeats, cysteine rich and spacers. ADAMTS-4 does not have further TS repeats. Proprotein convertase activity is required for the activation of ADAMTS 4 and 5; this enzyme allows for the removal the pro-domain to allow ACAN cleavage (Stanton et al., 2011). Cleavage sites for all the enzymes that degrade aggrecanare found in Figure 6.

Figure 6: Major proteolytic domains of ACAN



Cleavage domains residing in the aggrecan core protein. MMPs (M), ADAMTS (A), Calpain (C) and HtrA1 (H) and the protein cleavage sites are named (Struglics and Hansson, 2012, Struglics and Hansson, 2010, Chamberland et al., 2009).

In disease states where there is progressive loss of ACAN it is suggested that MMPs precede ADAMTS, temporal activation of the MMPs may due to the secretion and activation of the other enzymes. It is understood that MMP3 is required for the gradual truncation of the C-terminus of the G3 domain that is seen with age (Durigova et al., 2011). Cleavage of ACAN occurs in the IGD, N-terminus of the G1 and G2 domains are pathologically important and found in OA patients. It should be noted that ADAMTS mediated ACAN loss is not blocked by Tissue inhibitor of metalloproteinase (TIMP) -1 or TIMP-2 or MMP inhibitors (Troeberg and Nagase, 2012). In the KS domain the releases of GAG chains leading to loss of function are thought to be instrumental in the progression of OA. Another proteases thought to have clinical importance in cleaving ACAN in the IGD are the mammalian high-temperature requirement A family of serine proteases (HtrA) as it is found in abundance in adult cartilage and is elevated in human OA cartilage (Chamberland et al., 2009).

Furthermore, certain regions of ACAN are targets of Cathepsins B, D and K cleavage. Cathepsin K has been shown specificity to G1, CS2 and KS domains. This cleavage site produces CS fragments which interact with Cathepsin K to allow for collagenolytic activity (Struglics and Hansson, 2012). It is predicted that the Cathepsins along with the MMPs allow the normal turnover of ACAN but their role *in vivo* requires further investigations. Although evidence from the Chloe mice, in which the aggrecan IGD is resistant to MMP cleavage, develop normally with normal growth plate (Little et al., 2005).

In vitro experiments have identified M- calpain's abilities to cleave ACAN at the IGD, KS and CS1 domains. There are multiple calpain cleavage sites with the fragments appear in arthritic cartilage as well as suggested for normal turn over (Struglics and Hansson, 2010) However, their clinical importance requires further experimentation (Maehara et al., 2007).

1.3. Transcription

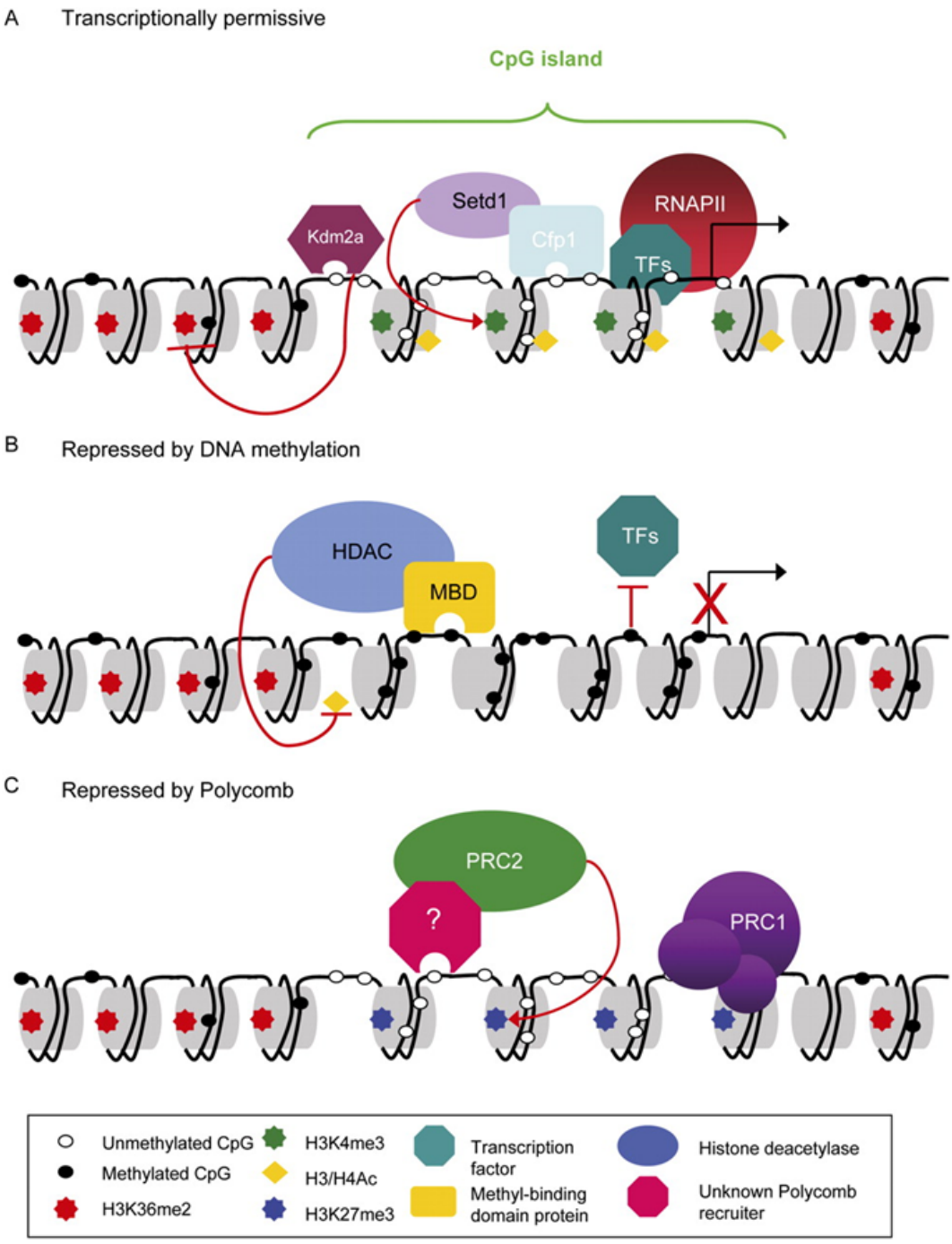
Transcription is the first step in eukaryotic gene expression, therefore its regulation underlines fundamental processes within the living organism such as differentiation and development. It is by no means static, but rather in development is episodic with a series of discontinuous bursts of gene activation and silencing occurring at the same time then an episode of relaxation while other genes are activated. Understanding how a gene is transcribed and the key regulators in this process provides avenues to understanding gene disorders and what happens in ageing. Additionally, provides targets for therapies and opportunities for development of tools to interrogate the signalling pathways and processes that define a cell.

1.3.1. The chromatin landscape

Each human cell contains around 2 metres of DNA, in the nucleus genes are encoded for by specific DNA sequences, which are organised around complex repeating unit of nucleoprotein (nucleosomes) structures collectively known as chromatins which need to be readily accessed during transcription and replication (Luger et al., 1997). The protein-DNA complex of two tight superhelical turns of approximately 146bp of DNA around disk

shaped proteins. These proteins are composed of eight histone molecules two copies of H2A, H2B, H3 and H4 (Soboleva et al., 2014). These provide not only densely packaged DNA, but regulate transcription by not only providing a physical barrier or bend DNA to restricted access for DNA binding elements such as TFs. The histones occur approximately every 200bp and have a “beads on string” structure with DNA coating around (Olins and Olins, 1974). TFs are DNA-recognition proteins that interact with cellular enzymes or allow the conformational change of the chromatin. The nucleosome is formed by histone chaperones disposition of the histones and chromatin remodellers. This is organised differently depending on the cell-type and specificity these are dependent on modifications that the nucleosome presents such as histone modifications and DNA methylation (Siggens and Ekwall, 2014). Transcription can be permissive, repressed by methylation or repressed by polycomb (Figure 7). Post-transcriptional modifications of histones can dictate the transcriptional landscape, such as the polycomb transcriptional-repression complexes. These are regulated by the PRC1 and PCR2 proteins known as PcG which are responsible for methylation of Histone 3 lysine 27 (H3K27) which results in a compact chromatin restricting access of TFs (Figure 7)(Di Croce and Helin, 2013). As cells differentiate the environment surrounding the DNA changes, the binding elements alters constantly which can dictate the activation of genes temporally.

Figure 7: The chromatin landscape



The transcriptional landscape is a dynamic state and at promoters where transcription is initiated the DNA can be exposed to the transcriptionally permissive by the work of chromatin remodellers or histone marks such as acetylation. These histones in which the DNA is wrapped around undergo posttranscriptional modifications that mark the

transcriptional state from repressed or permissive. Methylation results in a blocking of TFs reducing transcription efficiency whereas the actions of polycomb proteins results in a compact landscape repressing transcription (Deaton and Bird, 2011).

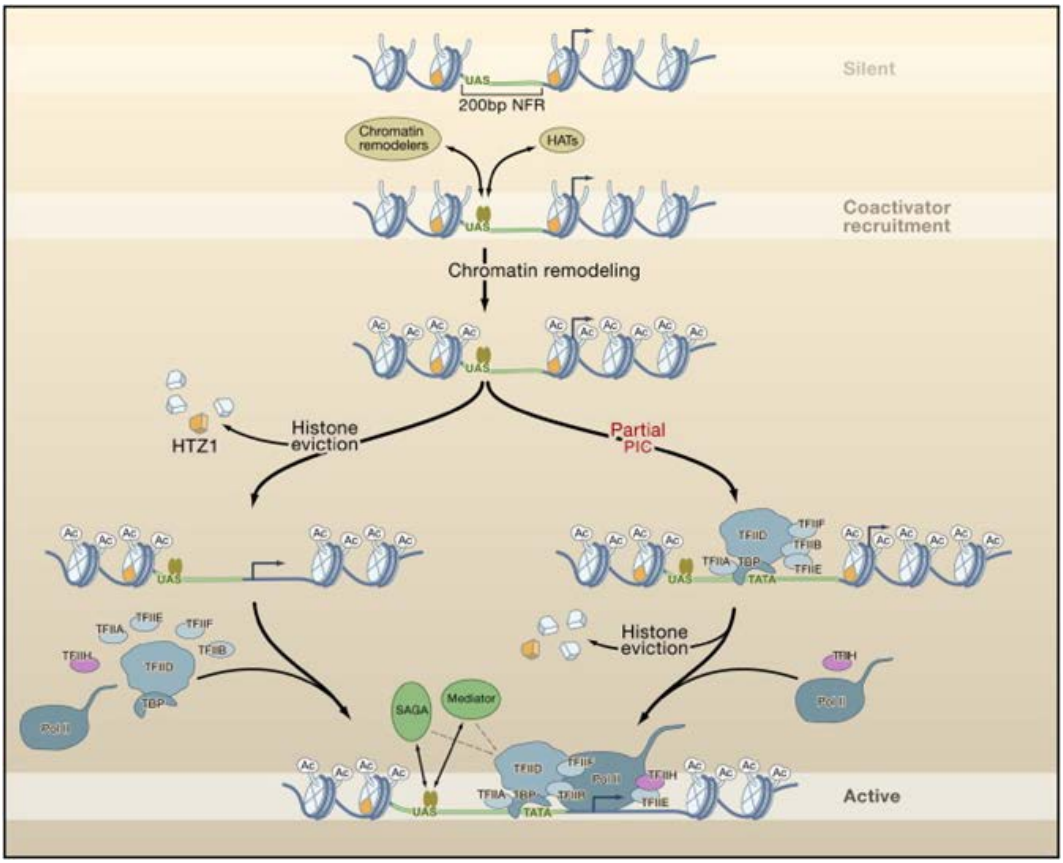
1.3.2. Mechanism of Transcription

In eukaryotic cells RNA polymerase II (Pol II) carries out major gene transcription (Roeder and Rutter, 1969). For transcription to occur the DNA must be first unwound to expose the DNA sequence to allow the binding of the transcription machinery (Boeger et al., 2003). The initial steps of transcription involve the recruitment of transcriptional activators that form the PIC (Figure 5) (Voss and Hager, 2014). The PIC at the promoter region bases before the TSS by the binding of activators and co-activators to specific sequences and completed by the recruitment of Pol II by protein- protein interactions (Murakami et al., 2013). The co-activator complexes act to strengthen the binding of activators and allow the DNA to be more accessible to the TFs. The histone is lost during transcription by the actions of ATP- dependent nucleosome-remodelling complexes. These promote the unravelling of the nucleosome by either moving or displacing the histones, this transient unwrapping of DNA, forming DNA loops or moving nucleosomes to allow better accessibility for TFs (Lorch et al., 2005, Lorch et al., 2010). Conversely, not all nucleosomes are removed; acetylated histones appear to accumulate during gene activation. Histones are not static, especially in actively transcribed genes in which histone dimers can be easily exchanged into and out of the nucleosome with the aid of chaperones. Additionally, the TFs binding sites of more active genes apparently positions in nucleosome free regions of DNA which are more exposed to the proteins in the nuclear environment.

Pol II is independently capable of unwinding DNA during synthesis of RNA yet alone it is unable to recognise a promoter and initiating transcription. To facilitate its actions a system comprising of evolutionary conserved proteins termed the general (or basal) TFs: TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH, are vital for basal transcription. The mediator, along with Pol II, is essential for transcription to start in almost all Pol II directed transcription and during mRNA elongation (Figure 8) (Kornberg, 2005, Takagi and

Kornberg, 2006). It facilitates PIC formation and bridges between protein activators at distant *cis*-regulatory elements such as enhancers and Pol II at the promoter (Kelleher et al., 1990, Robinson et al., 2016). Transcription starts at the assembly of nearly 60 proteins to form the transcription machinery (Conaway and Conaway, 1993). Pol II is positioned at the core promoter by a combination of TFIID, TFIIA and TFIIB. TFIIH allows the positioning of DNA into the PIC and RNA synthesis to occur. At this point the entire machinery may abort (abortive initiation) or escape (promoter escape) the promoter site, by the disassociation of TFIIB and the promoter sequence of the DNA, allowing transcription to be permitted (Liu et al., 2010). The "open" chromatin is a common feature of active transcription.

Figure 8: The transcriptional pathway

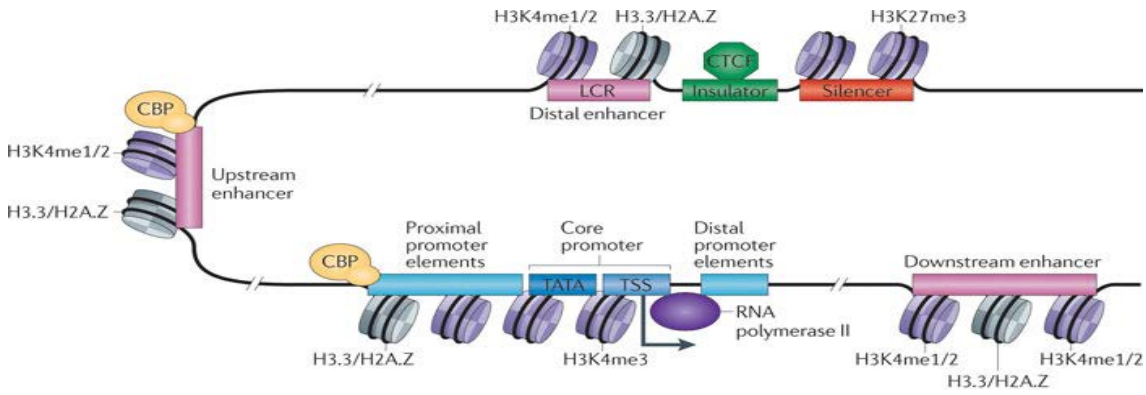


Transcription is initiated at the promoter by the recruitment of activators and chromatin remodelling enzymes. This action removes or modifies the histones that the DNA surround to allow the exposure of transcription factor binding sites and other interaction sites on the DNA. The preinitiation complex forms with the arrival of the general transcription factors and DNA polymerase II. The mediator then interacts with upstream or downstream elements to permit transcription (Li et al., 2007).

DNA enters the transcription machinery as a duplex, unwinds 3 bases before the active site of the machinery and sharply bends to allow the next nucleotide base to be flipped to face the active site. All 4 NTPs are found in the entry site of the machinery, they rotate to sample the DNA strand at the binding site. The addition of the NTP occurs by the motion of the trigger loop that is in contact with all moieties of the NTP. It swings like a trap door (Huang et al., 2010). The 2'-OH group that has been added to the RNA, the histidine promotes the flow of electrons during the nucleophilic attack of the 3'-OH acting as a proton donor triggering phosphodiester bond formation resulting in the positive selection of the complementary NTP. The wrong NTP creates a misalignment of the trigger loop therefore the network cannot form (Gnatt et al., 2001).

Elongation Pol II CTD undergoes two major phosphorylation events. Ser5 is phosphorylated by TFIIH in the 5' end of the open reading frame and Ser2 is phosphorylated by Ctk kinase as Pol II moves towards the 3' end. A bridge helix in Pol II bends to allow the translocation of the template strand while allowing the RNA strand and DNA to remain in close proximity (Gnatt et al., 2001).

Figure 9: The chromatin



The chromatin is composed of DNA wrapped around the histone octamers; these can permit or repress transcription. The promoter is situated around the transcription start site and is aided in the recruitment of activators and transcription factors by the proximal regions. Enhancers regulate transcription by interacting with the mediator at the promoter to facilitate transcription. Other elements such as insulators allow specific transcription and silencer to "turn off" transcription (Ong and Corces, 2011).

RNA is released from the transcription machinery by the effects of 3 protein loops- fork loop 1, rudder and lid. The Rudder and lid lie between the DNA and RNA with the lid in contact with the RNA and rudder with the DNA. The lid acts like a wedge keeping the two strands as separate entities. The fork loops stabilises the complex (Kornberg, 2007). However, the promoter requires further signals from *cis*-regulatory elements to allow transcription to occur. Such elements include enhancers, silencers, repressors and insulators (Figure 9).

1.3.3. Histone modifications

The mammalian genome is partitioned into two states, the euchromatin (meaning permissive of transcription) and heterochromatin (inhibitory of transcription), the heterochromatin is generally considered as silencing of the gene, particularly in development and is enforced by the Polycomb group and is able to switch which is noted in diseases such as cancer (Dejardin, 2015, Trojer and Reinberg, 2007). Both the

euchromatin and heterochromatin are littered with histone modifications. The 4 histones are composed of the site of active transcription is free of nucleosomes whereas the histones that surround these sites show evidence of post transcriptional modification. These modifications can be said to be markers of the activity states of transcription. Histone tails and globular domains are subjected to post-translational changes including methylation, acetylation (by histone acetyltransferase complexes), ubiquitination, ADP-ribosylation and phosphorylation of serines and threonines (Table 3, Table 4) (Lawrence et al., 2016).

Methylations can affect the ability of certain protein binding affinity to the chromatin. In particular methylation at lysine 4 (H3K4me) is associated with active transcription, a tri-methylation of histone 3 lysine 4 known as H3K4me3 is characteristic of transcriptional promoters due to the ability to recruit elements that can result in the formation of the PIC. Additionally, tri-methylation occurs around the TSS and appear to show the frequency of PolII that has travelled along this point. Conversely, it has little or no effect on elongation or Pol II processing but more like act like signalling centres. The nucleosome can be modified asymmetrically and symmetrically that appear to be of importance in development allowing certain genes to be poised for transcription especially in the case of methylation of lysine.

Acetylation of lysine increases the accessibility of the DNA wrapped around nucleosomes (Voss and Hager, 2014, Kouzarides, 2007). It is generally a transient event that is linked to active transcription and a permissive chromatin such as acetylation of lysine 27 of histone 4 (H4K27ac). The results of acetylation can reduce the positive charge of the histone resulting in a looser relationship between the DNA backbone and the histone tails allowing easy access for transcription factors (Soutoglou et al., 2000). As transcription is carried out there is dynamic acetylation and deacetylation and then resets the chromatin to reduce transcription getting out of hand. Additionally, heavily deacetylation of histones, along with cysteine methylation facilitates gene silencing.

These histone modifications have the potential to change the net charge of the DNA resulting in a weakness in the protein-DNA interactions that allow DNA to unravel or

histone to be displaced (Zhao et al., 2005). They may allow different proteins to interact by protein-protein interactions and therefore influence the higher order chromatin structure. What is clear is that they have no single function but many in a dynamic environment.

Table 3: Histone tail modification and their actions

Histone	Modification	Type	Effect
H2A	S1P	Phosphorylation	Chromatin assembly
	K4/5ac	Acetylation	Transcriptional activation
	K7ac	Acetylation	Transcriptional activation
	K119uq	Ubiquitination	Transcriptional repression
H2B	S33P	Phosphorylation	Transcriptional activation
	K11/12ac	Acetylation	Transcriptional activation
	K15/16ac	Acetylation	Transcriptional activation
	K123uq	Ubiquitination	Transcriptional activation
H3	K4me2	Methylation	Permissive euchromatin
	K4me3	Methylation	Transcriptional elongation
	R17me	Methylation	Transcriptional activation
	K27me3	Methylation	Transcriptional silencing; bivalent gene/ poising
	K36me3	Methylation	Transcriptional elongation
	K9ac	Acetylation	Histone deposition/ Transcriptional activation
	K14ac	Acetylation	DNA repair/ Transcriptional activation
	K18ac	Acetylation	DNA repair/ Transcriptional activation

	K27ac	Acetylation	Transcriptional activation
	T3P	Phosphorylation	Mitosis
	S10P	Phosphorylation	Mitosis/ Transcriptional activation
H4	R3me	Methylation	Transcriptional activation
	K20me1	Methylation	Transcriptional silencing
	K20me3	Methylation	Heterochromain
	K5ac	Acetylation	Histone deposition
	K8ac	Acetylation	Transcriptional activation/ DNA repair
	K12ac	Acetylation	Histone deposition
	S1P	Phosphorylation	Mitosis

Common histone tail modification on the N terminus and their effect on the nucleosome. K: Lysine, S: Serine, R: Arginine and T: Threonine. Table edited from (Lawrence et al., 2016) and (Karlic et al., 2010).

More recently novel histone modifications and cross-talk have been identified through the use of mass spectrometry (MS) and single molecule imaging (Shema et al., 2016). These include lysine crotonylation (the addition of the univalent radical crotonyl CH_3-CH=CH-CO^-), lysine butyrylation, lysine propionylation, lysine succinylation and lysine hydroxylation, incorporation of small ubiquitin-like modifiers (SUMOlyation) (Tan et al., 2011). Histone modifications can form though a combination of different modifications, histone modification interactions (Hall et al., 2009), this leads to a histone landscape that can vary depending genes and the reason for this diversity remains unclear (Rothbart and Strahl, 2014, Heintzman et al., 2009b, Henikoff and Shilatifard, 2011).

Table 4: Histone globular domain modifications

Histone	Modification	Type	Effect
H2A	K36ac	Acetylation	Unknown
	K99me	Methylation	Unknown

	K95me	Methylation	Unknown
	Q105me	Methylation	Reduce nucleosome occupancy
	K119ac	Acetylation	Unknown
	K199uq	Uniquitylation	Unknown
	K75me	Methylation	DNA binding
H2B	K40me	Methylation	DNA binding
	K82ac	Acetylation	Charge neutralisation
	R96me	Methylation	Unknown
	K105ac	Acetylation	Unknown
	K117ac	Acetylation	Unkown
H3	K122ac	Acetylation	Active transcription
	K115ac	Acetylation	Destabilisation of nucleosome
	K64ac	Acetylation	Transcriptional start/ mRNA generation
	K56ac	Acetylation	Transcriptional activation
	K56me3	Methylation	Repressive
	T118P	Phosphorylation	Nucleosome assembly
	K79me3	Methylation	Transcriptional start
	Y41P	Phosphorylation	Transcriptional activation

	R42me1	Methylation	Transcription activation
	K64me3	Methylation	Pericentromeric heterochromatin/ Nucleosome stability
H4	K91ac	Acetylation	Destabilisation of nucleosome
	K31ac	Acetylation	DNA binding
	S47P	Phosphorylation	DNA binding
	K79ac	Acetylation	Silencing
	R92me	Methylation	Unknown

Common histone globular domain modification on the N terminus and their effect on the nucleosome. K: Lysine, Q: Glutamine, S: Serine, R: Arginine, T: Threonine and Y: Tyrosine. Table edited from (Lawrence et al., 2016), (Tessarz and Kouzarides, 2014) and (Cosgrove et al., 2004).

Therefore, the histones give an indication of the transcriptional state and the structure of the chromatin which can be used to mark out regulatory elements in the genome and a small number of corresponding histone modification taken together can predict levels of gene expression (Karlic et al., 2010).

1.3.4. Variation of histones

Nucleosome-free region at active TSS are followed by a strongly positioned +1 nucleosome immediately downstream of the TSS that is independent from the class of promoter and is not constant throughout the cell cycle. The +1 nucleosome at active promoters is followed by uniform or phased arrangement of nucleosomes into the gene body; this is directional and apparently dependant on transcription since it is never found upstream of the promoter. Whereas a repressed gene lacks the phasing of nucleosomes and at the TSS is covered by histones (Lantermann et al., 2010, Valouev et al., 2011).

The H2A.Z histone variant is evolutionally conserved and critical for early development and gene expression regulation in embryonic stem cell differentiation is a variant of the H2A histone (Pandey and Dou, 2013, Subramanian et al., 2013, Hu et al., 2013). The function of the H2A.Z histone is associated with active transcription, since it is generally recognised at promoters or enhancers (Barski et al., 2007). When this histone is acetylated and ubiquitylated has shown to confer gene activation and repression (Sarcinella et al., 2007, Valdes-Mora et al., 2012). The incorporation of SUMOylated H2A.Z appears to destabilise the chromatin and initiate the DNA damage response (Kalocsay et al., 2009, Schwertman et al., 2016).

Other histone variants mark different events in transcription, such as the H3 variant H3.1 is found almost exclusively during DNA replication (Tagami et al., 2004) whereas H3.2 and non-allelic H3.3 mark interphase and nucleosome disassembly and reassembly that maintains chromatin structure during transcription (Ahmad and Henikoff, 2002).

These histone variants offer differences in the stability of the entire structure or exposes new sites for posttranslational modifications. For example during DNA double strand break repair the histone variant phosphorylated H2.AX (γH2.AX) cross talks with H3 acetylation the instant DNA damage occurs to open up the chromatin exposing binding sites for DNA repair proteins (Lee et al., 2010, Shema et al., 2016).

1.3.5. Promoters

The core promoter is considered as a structural and functional transcription regulator. It contains the minimal sequence required for the interaction of the transcription machinery to initiate transcription. The DNA binding motifs are recognised by the general transcription factors TFIID recognises the promoter sequence (it is composed of TATA-box binding protein (TBP) and TATA-box binding protein -associated factors). Core promoters are a collection of DNA binding motifs that can be found together or independently to bring about transcription at the transcription start site (TSS). These motifs include the classical TATA box (Thomas and Chiang, 2006), TFIIB recognition element (BREu and BREd) (Burke and Kadonaga, 1996), Initiator (Inr) (Smale and Baltimore, 1989), motif ten element

(MTE) (Lim et al., 2004), downstream core promoter element (DPE) (Burke and Kadonaga, 1997), Bridge, TCT, X core promoter element 1 (XCPE1) and downstream core element (DCE) (Kadonaga, 2012, Lewis et al., 2000). The TATA box is a classical site that is found at some eukaryotic promoters, it is located between 30-100 nucleotides before the TSS which the TATA box binding protein interacts with to partially unwind DNA and bend it 75-80 degrees towards the major groove (Thomas and Chiang, 2006). This allows distant parts of the DNA to be closer facilitating transcription factors to make contact. However, it is now know that fewer than 22% of promoter regions in eukaryotic cells contain TATA boxes (Gershenzon and Ioshikhes, 2005). Promoter can be distinguished into focus and dispersed. Focus is deemed by a single or a close cluster of promoter sequences where transcription is initiated from. Where dispersed is a collection of “weak” promoter sequences spanning 100 nucleotides and the most common in the genome (Juven-Gershon et al., 2008).

In vertebrates cytosine bases in the configuration of CG are highly underrepresented in the genome with the exception of CpG islands. Cytosine in genomic DNA can be methylated at the 5^{th} carbon to become 5-methylcytosine, these DNA methylation can act as a barrier for transcription attracting methylated DNA binding proteins that block binding sites (Siggens and Ekwall, 2014). The CpG regions are characterised by the CG repeats stretching on average 1kb with the dinucleotide, and unusually they are unmethylated. These are highly associated with promoter function and transcription start sites found in the 5’ region of genes (Spruijt and Vermeulen, 2014). Methylation of the 5’ of cytosine (5mC) can be read by 5mC readers that act, paradoxically, to regulate transcription on methylated cytosine, whereas normally methylation is normally a mark of gene silencing. Additional features of active promoters are the presence of histone modifications that prevent methylation events. The region immediately upstream of the core promoter is defined as the proximal promoter and is characterised by the multiple activator binding sites. The proximal region and the promoter tend to act synergistically as mutations in these regions dampen the ability of the core promoter (Hernandez-Garcia and Finer, 2014).

1.3.6. Insulators

Insulators play an important role in segmenting the chromosome and creating independent domains and restricting the effects of other *cis*-regulatory elements such as silencers and enhancer to their target gene only. There are two types of insulators: enhance-blocking and barrier (Sun and Elgin, 1999, Ali et al., 2016).

Enhancer blocking insulators are defined by their ability to interfere with the enhancer-promoter interactions when placed between the two. The mechanism of enhancer-blocking appears to be the ability to loop the DNA thus separating the two elements (Gaszner and Felsenfeld, 2006). In humans, this is done by the binding of the CCCTC-binding factor (CTCF) at the insulator sequence. CTCF is a zinc-finger protein works in two suggested models. CTCF binds to the chromatin fibre via the protein nucleophosmin creating a loop and keeping enhancers and promoters in different domains so they cannot interact (Yusufzai et al., 2004, Ali et al., 2016). This model would explain the short distances for enhancers but does not account for the long distance interaction of enhancers, this is where it has been shown that CTCF can reduce the acetylation of histones and the recruitment of pioneer and activating TFs such as CBP and p300 resulting in a blocking of PolII (Zhao and Dean, 2004). The way CTCF affects transcription may be dependent on the context and the strength/ activity of the enhancer in which it insulates (Varma et al., 2015).

Barrier insulators protect against position effects variegation, where a gene in a lightly packed chromatin area is juxtaposed with a heterochromatin by rearrangement or transposition therefore silencing the region (Elgin and Reuter, 2013). Barrier insulators operate independently from the CTCF protein but rely on a heterodimer of Upstream transcription factor 1 and 2 (USF1/2) that recruit histone modifying enzymes to the region such as histone methyltransferases and histone acetyltransferases to deposit high level of transcription repressing the heterochromatin protein 1 (HP1) and protects the histones from H3K9 methylation thus preventing the spread of the heterochromatin (Ghirlando et al., 2012, Gaszner and Felsenfeld, 2006).

1.3.7. Silencers/repressors

Silencers prevent gene expression and they are sequence specific elements that allow the binding of transcriptional repressors. There are three models that suggests how a silencer operates, in the first they silence transcription by interfering with PolII or pioneer factors (mainly TFIIB) from binding by promoting local chromatin condensation. The second is by the formation of transcriptional block by competition of a binding site, disrupting quenching or inhibiting activating TFs, forming a heterochromatin (Johnson et al., 2013, Thurman et al., 2012). For example the recruitment of polycomb complexes PRC1 and PRC2 to target sites condenses the chromatin forming a physical barrier for transcription (Chopra et al., 2012, Johnson et al., 2013). The final by anti-looping between enhancers and promoter (Li and Arnosti, 2011, Chopra et al., 2012, Chen and Widom, 2005). In *Drosophila* it has been observed that *Snail* acts a repressor by binding at silencers and blocking enhancer- promoter looping and by inhibiting release of PolII from the transcription machinery (Chopra et al., 2012, Bothma et al., 2011). Silencers can be found in the proximal promoter, within enhancers or as standalone elements. For instance the *Pax8* proximal promoter contains as a silencer, the *Pax8* silencer drives tissue specific repression to allow expression in a tissue specific manner, such as the tail bud in *Xenopus* (Ochi et al., 2012). This silencer contains repressor motifs for RE1- silencing transcription factor (REST) and the homobox family Nkx which recruits histone methlytranferases (Muhr et al., 2001).

1.3.8. Enhancers

Enhancers are clusters of transcription factor binding motifs, those close to the promoter site are recognised as proximal promoters, first identified in viral genomes they were defined simply as DNA fragments that are located outside of the core promoter that increases the transcription of another gene (Banerji et al., 1981). These can regulate transcription by allowing distant transcription factors to interact with activators and co-activators at the promoter or by direct contact with the promoter of the target gene (Yao et al., 2015). They play an active role in transcriptional bursting, periods of time where many transcripts are formed by release of PolII from active promoters then a refractory

period therefore instrumental in developmental gene activation. Areas of active enhancers are void of nucleosomes making the DNA regions accessible to the transcription machinary. Furthermore, they are marked by certain histone modifications to show a permissive chromatin such as acetylation of the 3rd histone at the 27th lysine (H3K27ac) and H3K4me1 (Heintzman et al., 2009a, Heintzman et al., 2007). They are able act independently of distance and orientation in relation to their target gene and are generally highly conserved in evolution. Enhancers' activation coincides with DNase1 hypersensitivity (Lelli et al., 2012). Their suggested mode of action is by bending the DNA in loops to interact with their promoters even from a distance or by direct interactions (Tolhuis et al., 2002, Deng et al., 2012).

1.3.9. Classification of enhancers

1.3.9.1. Locus control regions

The locus control region (LCRs) was first described in the β- haemoglobin gene, they are group of DNA elements that can regulate an entire locus or gene cluster (Schubeler et al., 2001, Talbot et al., 1989). Defined by the ability to drive tissue specific and physiological expression of a linked transgene in a position-dependant and copy number dependant manner from a distance (Li et al., 2002). LCRs do not just encompass enhancers but include silencers, insulators, scaffold attachment regions and nuclear matrix binding sites that are regulated by TFs, coactivator and chromatin modifiers. They act as chromatin hubs that allow an open chromatin to loop to exclude or bring regions closer or further away to activate or repress transcription (Maston et al., 2006, Krivega and Dean, 2016). LCRs are uncommon in the genome and many genes do not contain LCRs, but studies have identified the human CD2 LCR is able to control cell lineage, suggesting the regulation of gene transcription may have evolved to encompass different mechanism that may be due to importance of the gene (Menzel et al., 2011). Furthermore, LCRs have been identified in the Th2 cytokine family IL-4, IL-5 and IL-3 (Lee et al., 2003) as well as in the human growth hormone multigene cluster LCR is able to form long range looping to form a hub to regulate the cluster in different tissues such as the placenta or pituitary (Tsai et al., 2016).

1.3.9.2. Stretch enhancers

Stretch/ stitched enhancers (SEs) are clusters of enhancers, sharing the same selection criteria as normal enhancers, spanning greater than 3kb with no gap and are tissue specific with an overlap with the LCRs. Cell- specific gene activation appears to reside in SEs with genetic traits in genome wide association studies and are housed within these regions (Parker et al., 2013, Quang et al., 2015). They are thought to provide flexibility, precision and diversity in expression of important genes and resemble super-enhancers but without a high affinity to a master TF (Quang et al., 2015).

1.3.9.3. Shadow enhancers

Nature tends to have a backup for stochastic, genetic and environmental perturbations, this is known as canalisation (Siegal and Bergman, 2002). On a transcriptional level it has emerged that developmentally important genes have functional redundancy by using two enhancers that drive transcription with overlapping functions, normally found with the normal enhancer within the first intron and the second, coined the “shadow” enhancer, and is found upstream of the promoter. They also confer a more precise expression of the gene (Hong et al., 2008, Hobert, 2010, Perry et al., 2010, Frankel et al., 2010, Metzger et al., 2015). Shadow enhancers have been mainly described in the *Drosophila* studies into patterning where if one enhancer is removed, in optimum conditions there is no effects on the gene being transcribed, however, in extreme conditions such as high temperatures the loss of the enhancer is recovered by the second enhancer (Perry et al., 2010, Frankel et al., 2010, Staller et al., 2015). It must be noted that functionally redundant enhancers appear widespread in the genome suggesting that there is a level of transcriptional regulation we do not yet comprehend (Cannavò et al., 2016).

1.3.9.4. Super-enhancers

Recently the emergence of a new class of enhancers entered the field, termed Super-enhancers. They are described as a group of putative enhancers in close genomic proximity with unusually high mediator (Med1) binding (Whyte et al., 2013, Loven et al., 2013, Hnisz et al., 2013, Hnisz et al., 2015). The original selection criteria for super-

enhancers were regions of the genome that were enriched in the master TFs Oct4, Sox2 and Nanog based on ChIP-seq within 12.5kb of each other were known as stitched enhancers. These stitched enhancers were then ranked by Med1 binding with no functional assessment, the enhancers not enriched for Med1 were termed “normal” enhancers (Whyte et al., 2013, Pott and Lieb, 2014). However, the biological rational for the term super-enhancer is only based on the enrichment of histone marks and transcription factor. What remains to be determined is their functionality. Other groups have looked to define super-enhancers as binding to the master TFs for each cell type such as Sox9 for chondrocytes (Loven et al., 2013, Liu and Lefebvre, 2015). However, it is becoming clear without more functional studies and better characterisation there appears to be no evolutionary benefit or any difference between a super-enhancer as opposed to a cluster of normal enhancers, for example, within super-enhancers there are some single enhancer elements that are able to drive expression in the absence of the other enhancers in the cluster and it is also known that enhancers themselves can act co-operatively (Whyte et al., 2013, Pott and Lieb, 2014, Carey et al., 1990).

Although the differentiation between enhancers and super-enhancers is weak it is certain, that the genome is highly enriched for enhancers that drive high level expression of a gene. Their importance is undisputed, as a single enhancer in these clusters can delay the time specific activation of a gene, the lineage fate of an entire cell, cell identity and in cancer allows heightened expression of oncogenes (Hnisz et al., 2013, Hnisz et al., 2015, Huang et al., 2016, Proudhon et al., 2016).

1.3.9.5. Temporarily phenotypic enhancers

Clustered regularly interspaced short palindromic repeats (CRISPR) and its CRISPR associated protein (Cas9) has revolutionised genome editing, using bacterial defence to remove foreign DNA, it has been modified to target regions of the genome to delete, insert, activate and repress regions with high precision (Carroll et al., 2016, Chu et al., 2015b, Dow et al., 2015, Ran et al., 2015, Sander and Joung, 2014, Shalem et al., 2015, Wu et al., 2013, Wu et al., 2015, Yin et al., 2014, Hwang et al., 2013, Thakore et al., 2015). CRISPR/cas9 has been utilised to identify on a large scale enhancers and to delineate the

higher chromatin order (Dominguez et al., 2016, Gilbert et al., 2013, Huang et al., 2016, Korkmaz et al., 2016, Li et al., 2014, Meyer et al., 2015, Miura et al., 2015, Nissim et al., 2014, Pefanis et al., 2015). A recent development in enhancer biology is the identification of a class of enhancers termed "Temporarily phenotypic enhancers. This work is in its infancy and in dire need for further investigation it demonstrates the importance of technology guiding better understanding of the eukaryotic gene regulation. The identification of this class of enhancers arise from a screen for *cis*-acting elements of the POU5F1 gene in embryonic stem cells using 1964 RNA guides for CRISPR of the topological associated domain (TAD) identified 174 candidates. Introduction of mutations into 2 of the elements (DHS_65 and DHS_108) and using the technique of local condition phasing of an enhancer green fluorescence protein (eGFP) to detect the changes in the expression of the promoter due to these mutations, identified that the cells were able to recover from a dip in eGFP expression after loss of the enhancer elements (Diao et al., 2016, Selvaraj et al., 2013). The group termed these Temp enhancers, and describes them as a non-canonical regulatory element that resides in an open chromatin, with low/weak level of activity *in vitro* activity, distal from their cognate promoter but their loss leads to a transient loss of activity at the promoter (Diao et al., 2016).

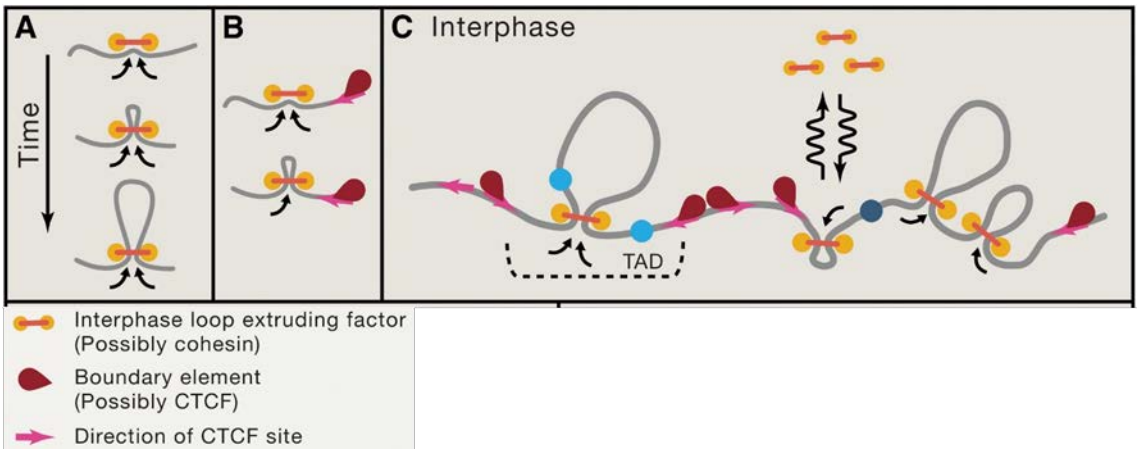
1.4. The 3D genome: looping in and looping out

Enhancers allow communication in the chromosome by interacting with promoters or direct contact with a corresponding loci, however, direct physical association does not account for all chromosome cross-talk as stochastic contact between proteins only occur when they are in close proximity. There are other proposed modes of communication such as diffusion of TFs, chromosome organisation and long range interactions. This is true for the epigenetic X-chromosome inactivation that communicates in more than one form. Inactivation is controlled by the long noncoding RNA *Xist* and its regulatory complex X-inactivation centre (Xic). *Xist* forms at the cis-acting elements in the *Xic* and spreads the signal to inactivating proteins to gradually silence the X-chromosome, this also requires initial inter-chromosome communication by transient contact to initiate signals by TFs to

allow long range chromatin interactions and eventual chromatin spreading and silencing and (Galupa and Heard, 2015, Chu et al., 2015a).

Genome compartmentalisation into megabases that reduces the distance between regions of the chromosome so during the cell cycle these regions can interact within one cycle. Active regions interact with active regions and inactive with inactive, this allows genes to be highly expressed during the cell cycle. How these boundaries regulate gene activation is still poorly understood, but it is thought that looping of the chromatin facilitates interactions, or excludes regions to silence the region and it is believed that enhancers allows the looping of these regions (Figure 10) (Dekker and Mirny, 2016, Chopra et al., 2012, Deng et al., 2012, Fukaya et al., 2016, Goloborodko et al., 2016a, Goloborodko et al., 2016b). In a study that challenges the looping mechanism, the *Drosophila* enhancers' *hunchback* proximal enhancer and *even-skipped (eve)* stripe 2 enhancer activity was visualised by labelling their target promoters MS2-YFP that contains the *sna* shadow enhancer and promoter and the MCP-GFP (Fukaya et al., 2016). The study showed that weak enhancers do not promote strong transcriptional bursts compared to strong enhancers and that bursts of transcription were coordinated at each promoter that share the two enhancers rather than enhancers alternating promoters, and forming loops to activate them, this suggests that promoters and enhancers not only forms loops but are directly linked or that promoters are linked to each other and that transcriptional activation is coordinated (Fukaya et al., 2016). With advancing technology in chromosome capture and imaging of chromosomes, it may reveal the role of enhancers and their role of interacting with promoters.

Figure 10: Looping interactions during the cell cycle that regulate gene activation



Loop extrusion (A) creates distances between regions of the chromatin to prevent interaction while bringing regions of the DNA to closer proximity which can be blocked by insulator regions or boundary elements such as CTCF (B). Looping allows the formation of domains in which transcription can be regulated more tightly and can allow long range interactions (Dekker and Mirny, 2016).

1.5. Transcriptional regulation of ACAN

1.5.1. The promoter

Unfortunately there is little known of the regulation of the ACAN gene despite its essential role in cartilage. In regards to the transcriptional regulation of the *Acan* a screen of rat, mouse and human aggrecan gene showed that it lacks classical TATA box binding sites, which is the location the TBP binds to initiate transcription or a CCAAT box (Doege et al., 1991, Doege et al., 1994, Watanabe et al., 1995, Pirok et al., 1997). However, at the predicted position there is a TATG site and multiple AP-2, AP-4 and SP-1 binding sites that can mediate the transcription start of TATA-less promoters (Valhmu, 1998, Pirok et al., 1997). Further studies have linked other factors that can influence the transcription of aggrecan such as TGF-β signalling (Radons et al., 2006), HIF-2α (hypoxic conditions) (Lafont et al., 2007) and hydrostatic pressure (Neidlinger-Wilke et al., 2006). Recently, *Acan* has been shown to be a target of SHOX proteins along with Sox9, which are able to

upregulate *Acan* mRNA and protein levels (Aza-Carmona et al., 2014, Aza-Carmona et al., 2011, Tiecke et al., 2006).

1.5.2. Silencer

When determining the regulatory sequences in *Acan* a 1.8kb sequence was gradually reduced in pGl2 luciferase reporter assays that were transfected into 14 day chick sternal chondrocytes (Pirok et al., 1997). Two silencer sequences were identified in the 5' UTR of the chick sequence. The first sequence located between -873 and -721, contained the CACCTCC (CIIS2) sequence which was shown to repress *Col2a1* promoter (Savagner et al., 1990). The second sequence ACCCTCTCT (CIIS1) was found at position 127 is identical to the *Col2a1* sequence and location. The removal of these two sequences removes the tissue specificity of the *Acan* promoter expression, suggesting that the silencer acts in different tissue types to down regulate *Acan* expression (Pirok et al., 1997). The *Col2a1* silencers are regulated by mouse *Snail* family *Snail and Slug* which acts as transcriptional repressors interacting with FGF receptors to down regulate ACAN and Col2a1 in pre-hypertrophic chondrocytes (Seki et al., 2003, Hong et al., 2009).

1.5.3. The untranslated region

The 5' and 3' UTR of genes tend to play critical roles in regulating transcription, the *Acan* UTRs are no different. To determine the role of the UTR the 5' and 3' regions were cloned to the pGEM-*luc* reporter and upstream of the CMV promoter and then transfected in to bovine articular chondrocytes. The 5' region alone significantly down regulated the luciferase activity by 92% (Valhmu, 1998). However, when the *Acan* promoter replaced the CMV promoter and the 5' UTR was placed directly upstream the expression increased by 7.7 fold when compared to promoter alone. Whereas the 3'UTR downregulated the *Acan* promoter expression by 68%whereas and together the UTRs reduced expression by 45% (Valhmu, 1998). It is noted that sequential removal of the promoter had little effect if the 5'UTR is present but removal of the 5' UTR reduced expression from between 85% to 100% suggesting this region can act as a promoter for *Acan* (Valhmu, 1998).

1.5.4. Enhancer elements in the aggrecan gene

In terms of enhancer elements of *Acan* recent work has attempted to delineate the regulatory elements of the gene. Doege and colleagues showed that a 4.6kb segment 12kb upstream of the transcription start site in the rat *Acan* was able to activate ACAN expression in chondrocyte cells. Additionally, this region weakly promoted ACAN expression in transgenic mice embryos E14.5, in a cartilage specific manner, but only in hypertrophic chondrocytes (Doege et al., 2002). This segment does not however drive ACAN expression in intervertebral discs and distal digits in development, suggesting other regions in the gene may cooperate in ACAN expression *in vivo*. One verified highly conserved enhancer of *Acan*, approximately -10kb in the mouse gene found in the 4.6kb segment previously analysed, was first described by Han and Lefebvre (2008), and with which the Sox trio interacts to allow the normal function of the enhancer. They noted Sox9 required the interactions of the L-Sox5 and Sox6 to bind to the enhancer. This enhancer element could drive ACAN expression in both transfected cells and in vivo in both adult and embryonic tissue in a specific manner typical of ACAN expression (Han and Lefebvre, 2008). SHOX and SHOX2 interact with Sox9, L-Sox5 and Sox6 to activate the -10kb *Acan* enhancer shown by yeast two-hybrid assay and luciferase expression in human bone osteosarcoma epithelial cells. These TFs collectively could allow activation of ACAN in limb development. Hu and colleagues utilised ENCODE data and conservation in the human aggrecan gene to locate possible enhancer motifs. They identified 11 regions capable of regulating ACAN expression in zebrafish. In these 11 regions, one was homologous to the known enhancer region. However, in zebrafish the expression was apparent in skeletal muscle which diverges from expression described by Han and Lefebvre. 8 of the 11 regions regulated cartilage specific expression at varying time points in development with the remaining regions only expressing in certain cartilage cells. 6 of the 11 regions were conserved at the chicken locus. Upon transfection of the chicken homologous sequence into zebrafish the results showed that 5 of the sequences presented with conserved expression. Of these regions, two had predicted sites of Sox binding, two with multiple transcription factor binding sites and one with none predicted (Hu et al., 2012). Their data suggests that *Acan* has multiple enhancer regions which may

regulate ACAN expression at different stages of development or they may form part of a redundant regulatory system. Ikeda and colleagues reanalysed the Han and Lefebvre (2008) regions. They showed that in the region -10kb from the *Acan* TSS (which was named A1) there is another element capable of expressing ACAN strongly in transiently transfected chondrocytes (Ikeda et al., 2014). However, they failed to show its expression *in vivo*. This data points to a greater transcriptional control of *Acan* compared to other ECM based genes.

2. AIMS AND OBJECTIVES

The aim of this body of work is to identify and verify whether *Acan* expression is driven by other enhancer elements. The revelation of TFs that bind to the core sequences and regulate the chondrocyte phenotypic expression of these regions or to establish if there is temporal expression of different TFs which allow differential activation of *Acan* would allow an appreciation of the mechanism of transcription. This is because it appears to be more complex for this gene based on current knowledge.

The transcriptional regulation of aggrecan in adult articular cartilage remain vague. It is uncertain if developmental expression of ACAN and adulthood expression of ACAN is maintained by different *cis*- acting elements or if there is a balance of other contributing factors. Therefore, the establishment of any enhancers identified is specific to adulthood would be essential for the understanding of maintenance of adulthood expression of ACAN; if particular enhancers interact with different transcription factors during this age or if there is redundancy in the system, would also be a critical understanding to the transcriptional environment of ACAN.

It is known that ACAN and the mRNA of *Acan* are found in other non- cartilaginous tissues, but the drivers of this expression are poorly understood. Different enhancers can allow tissue, or even cell type, specific manifestation of certain genes or the possibility that the enhancers of one tissue type represses that expression in other tissues. Therefore, the identification of other enhancers that define tissue expression can set the ground work for future studies into non-cartilage degenerative disease involving ACAN.

An appreciation of the transcriptional regulation of aggrecan can give us the basic understanding of normal systems within joints and the possible pathways that deteriorate in disease states such as osteoarthritis, where the loss of ACAN is one of the major events. Once identified the key regulatory regions of *Acan* it is in hope that these sites can be targets for cell-specific treatments of such diseases as enhanceropathies are being established as modes for disease progression.

3. MATERIALS AND METHODS

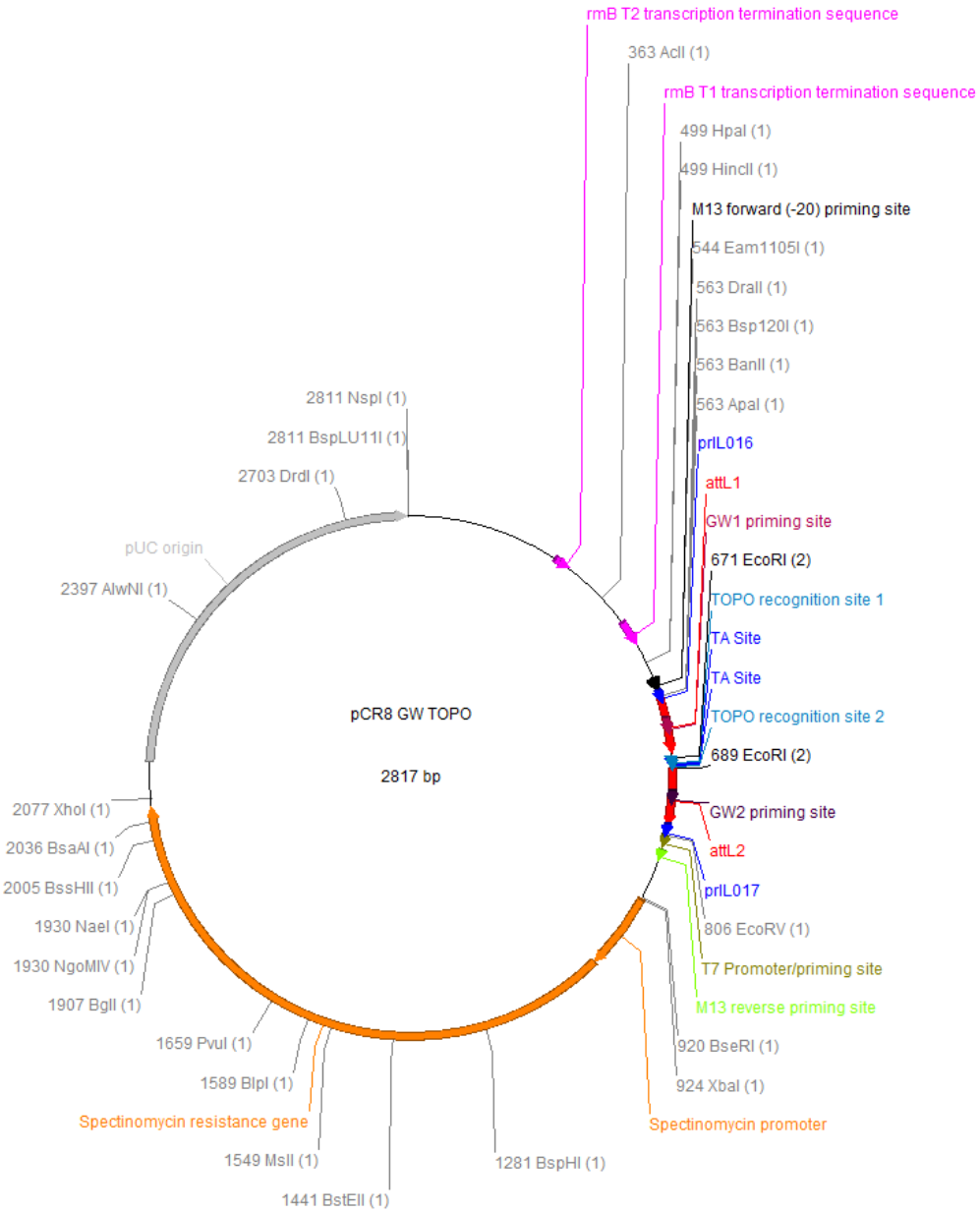
3.1. Competent Bacterial Strains

For general cloning the competent bacteria used for were the One Shot® TOP10 Chemically Competent *Escherchia coli* (*E.coli*) (ThermoFisher Scientific, C404010) or in-house propagated TOP10s. For cloning the pGl3 luciferase vectors Subcloning Efficiency™ DH5α™ (ThermoFisher Scientific, 18265-017), Single-Use JM109 (Promega, L2005) and JM109 propagated in-house were used.

3.1 Cloning vectors

3.1.1. Invitrogen™ Gateway™ Cloning

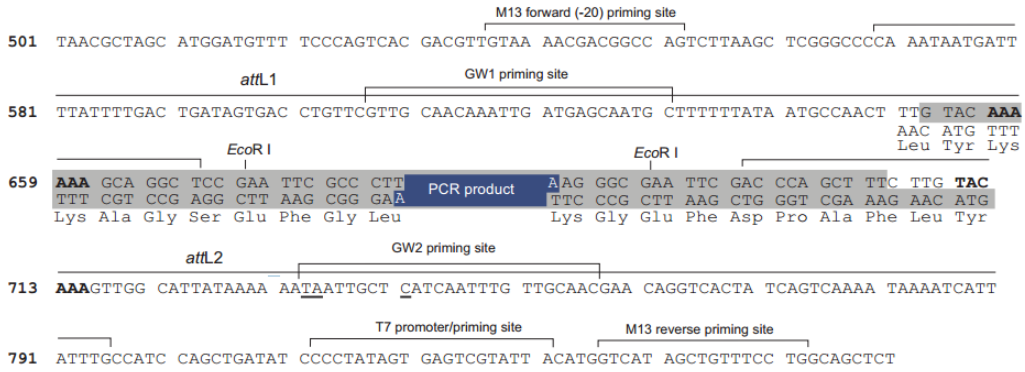
Figure 11: Schematic pCR8 GW vector



3.1.2. pCR™8/GW/TOPO® TA

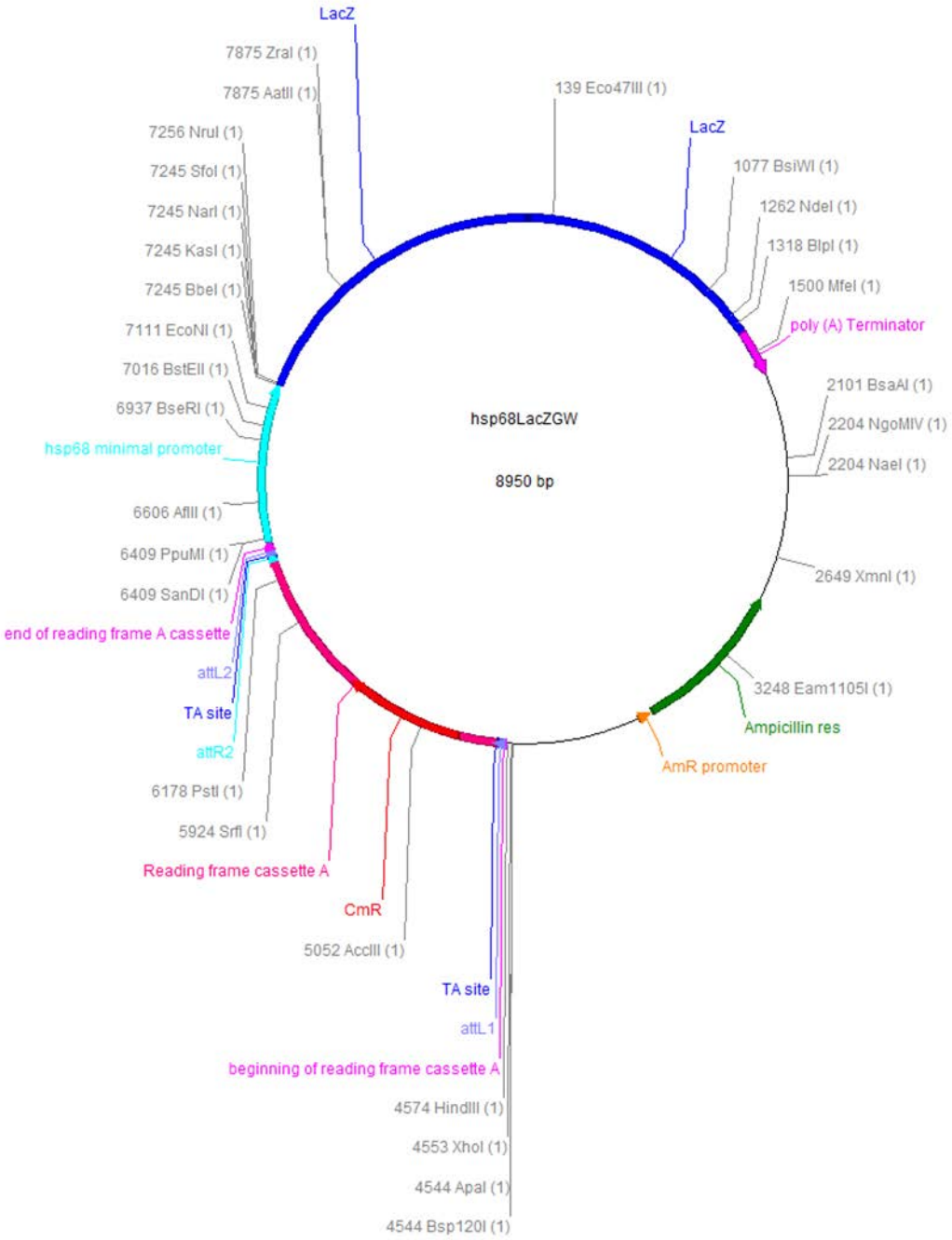
The pCR™8/GW/TOPO® vector was used in TA based cloning to place PCR products into an intermediate ready for further cloning, it creates an entry vector for Gateway® cloning. The TOPO® utilises a single 3' thymidine to take *Taq* polymerase amplified products. It benefits from Topoisomerase I from *Vaccinia* virus covalently bound to the ends of the linear vector at the TA site that binds to "CCCTT" sites on DNA to break the phosphodiester bonds on one strand of DNA and releases energy that is used to for another bond between 3' phosphate of the cleaved strand and the Tyr-274 residue on the Topoisomerase I. The *att*L sites that flank the inserted PCR product allows the flipping of the insert into a destination vector.

Figure 12: TOPO insert site



3.1.3. HSP68/LacZ/GW

Figure 13: HSP68/*LacZ*/GW vector schematic



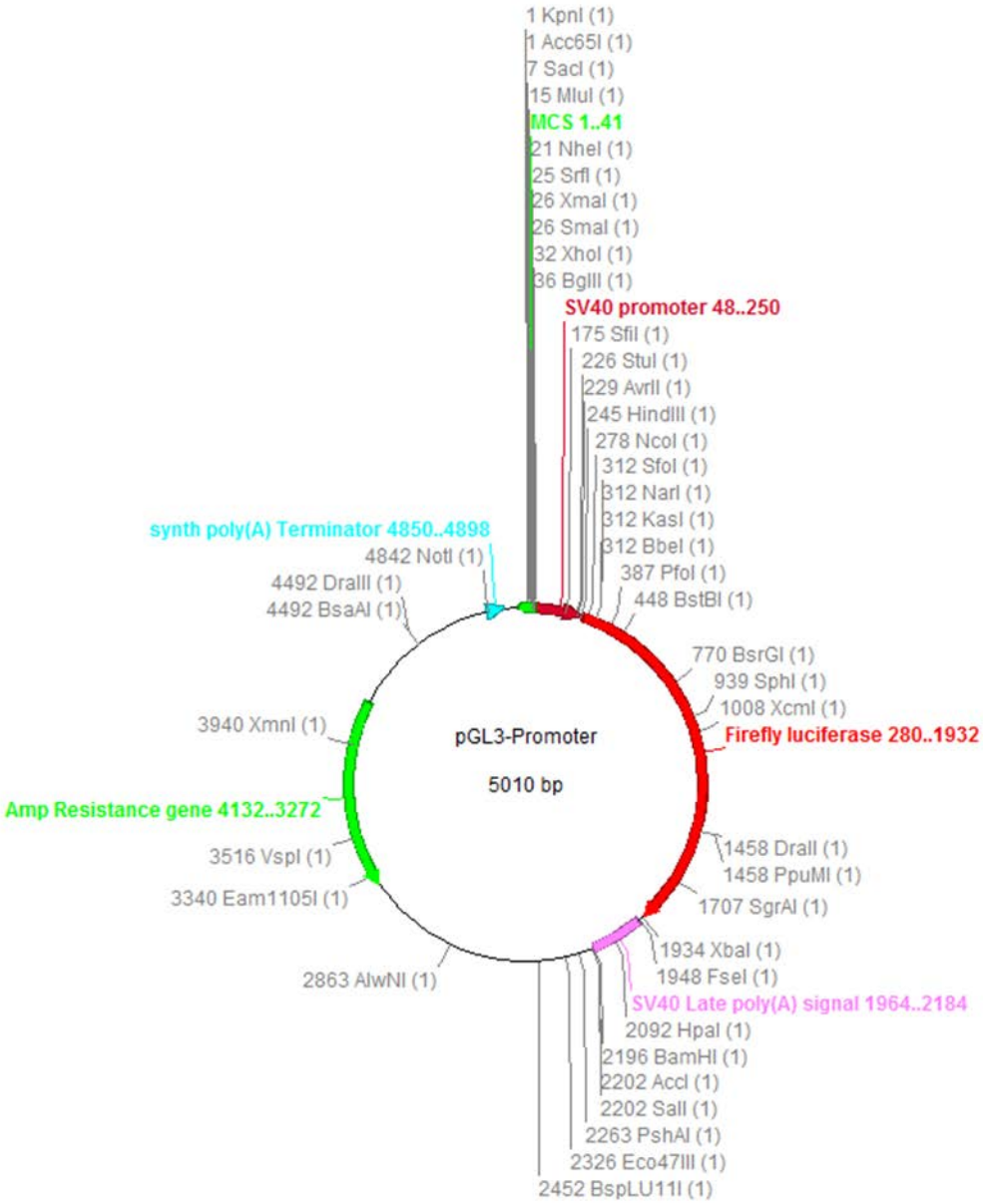
The HSP68/*LacZ*/GW was a gift from Nadav Ahituv (Pennacchio et al., 2006), it is used to generate the transgenic mice used in this study. It contains a Gateway® cassette fused the heatshock protein 68 minimum promoter (HSP68) (-664 to +113bp/ heatshock protein family A member 1A) and *LacZ* gene (Kothary et al., 1989) which has no constitutive expression in transgenic mice. The Gateway® cassette allows the vector to be the destination vector when the entry vector is introduced in the presence of the bacteriophage lambda recombination protein Integrase and Excisionase found in the LR Clonase mix (Invitrogen).

3.2. Luciferase based vectors

3.2.1. pGl3- Promoter

The pGl3-Promoter has MSC to allow insertion of genes/ sequences of interest. It contains a modified *Photinus pyralis firefly* luciferase and synthetic poly(A). It houses a Simian virus 40 (SV40) promoter with MCS upstream and downstream to assess the function of DNA elements. This was used to characterise the expression of the *Acan* enhancers in cells in the absence of the endogenous promoter

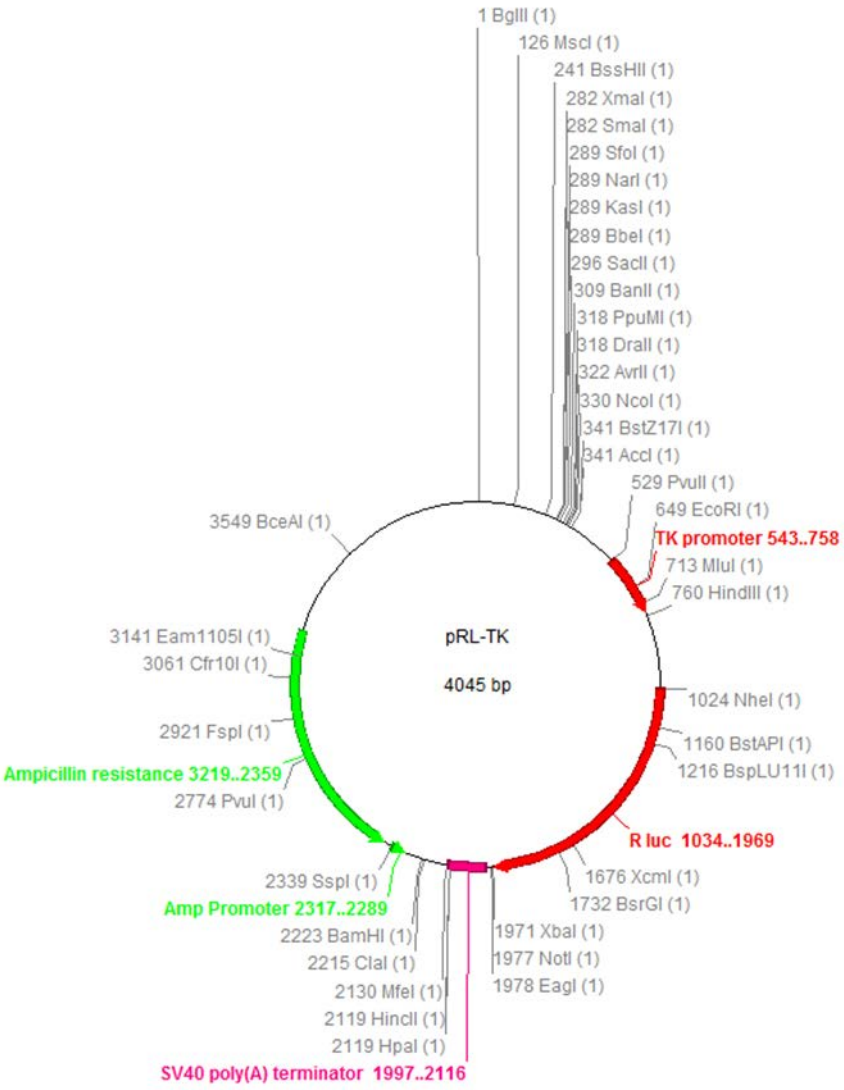
Figure 14: pGl3-promoter vector



3.2.2. pRL-TK

The pRL-TK contains a *Renilla* luciferase and is used as an internal control reporter to normalise for transfection efficiency. The herpes simplex virus thymidine kinase (TK) promoter luciferase was chosen due to its low level of expression when transfected into mammalian cells.

Figure 15: pRL-TK vector



3.3. Plasmid preparation

3.3.1. Making TOP10 competent cells for general cloning

3.3.1.1. Recipes

3.3.1.1.1. TFBI

Reagent	For 500mL
30mM Potassium acetate (KOAc)	1.45g
50mM Manganese (II) chloride ($MnCl_2$)*	4.9g
0.1M Rubidium chloride (RbCl)	6g
10mM Calcium chloride ($CaCl_2$)	0.735g
15% Glycerol	75mL of 100% glycerol
0.2M Acetic acid	Use to pH to 5.2
dH_2O	To 500mL

**$MnCl_2$ was added last as it precipitates otherwise*

TFBI was filter sterilised using a 0.2µm filter and store in 45mL aliquots at -20°C.

3.3.1.1.2. TFBII

Reagent	For 200mL
10mM MOPS sodium salt pH7	20mL
75mM Calcium chloride ($CaCl_2$)	2.2g
10mM Rubidium chloride (RbCl)	0.24g
15% Glycerol	30mL

TFBII was filter sterilised using a 0.2µm filter and stored in 12mL aliquots in the dark at -20°C.

Chemically competent *E.coli* used for propagation of vector constructs were grown in-house from a small amount of One Shot® TOP10 Chemically Competent *E.coli*. A portion of the original bacterial stock was collected using a sterile inoculation loop and streaked onto

an antibiotic free Luria-Bertani with Lennox L agar (LB-agar) plate. The plate was subsequently incubated overnight at 37°C. The following day 4mL of Luria-Bertani broth (LB-broth) was inoculated by a single colony from the LB agar plate and incubated for 16 hours at 37°C shaking at 150rpm to create a starter culture. The starter culture was transferred to 2x 150mL of LB broth and grown in at 37°C shaking incubator (225rpm) for 2-4hours until the optical density reached 0.3-0.4 at wavelength 600nm. The culture was then centrifuged at 8,000g for 10 minutes to create a bacterial pellet and the supernatant aspirated and disposed of. The pellet was resuspended in 10mL of ice cold TFBI on ice by trituration and incubated on ice for 15 minutes and then centrifuged at 8,000g for 5 minutes. The supernatant was aspirated and disposed of and the remaining pellet was resuspended slowly in 500 µL ice cold TFBII on ice by pipetting. The resultant chemically competent bacteria was pooled and diluted in 8mL of ice cold TFBII, inverted to mix, aliquoted, snap frozen in liquid nitrogen and stored at -80°C until use. To check for errors a sample was plated on LB agar plate which contained 100µg/mL of ampicillin to ensure no growth occurred after overnight incubation. To check the transformation competency, pUC19 were transformed in parallel with the new stock compared to the old stock in a dilution series to ensure similar colony-forming units (CFUs).

3.3.2. Preparation of JM109 or DH5α *E.coli* competent cells

3.3.2.1. Solutions

3.3.2.1.1. 60mM $CaCl_2$

Reagent	**For 500mL**
$CaCl_2$ (Sigma)	4.4g
1,4-Piperazinediethanesulfonic acid (PIPES)	1.73g
Glycerol	70mL
ddH_2O	To 500mL

Adjusted to pH7 with NaOH, filter sterilised using 0.2µM syringe filter and stored at 4°C.

A stock of JM109 or DH5α *E.coli* was streaked onto LB agar plate with no antibiotics and incubated overnight at 37°. A single colony was selected from the plate and incubated in 5mL of LB broth with no antibiotics at 37°C shaking at 225rpm to create a starter culture. 2 x 250mL LB broth with no antibiotics were inoculated with 2.5mL of starter culture and then incubated at 37°C until OD at 600 was between 0.3 -0.4. The 250mL cultures were divided into 50mL Ultra High performance (up to 20,000g) centrifuge polypropylene tubes and centrifuged at 1600g to produce a pellet. The supernatant was discarded and the bacterial pellet was pooled by suspending gently on ice in 40mL of ice cold 60mM $CaCl_2$. The suspension was centrifuged at 4°C at 1100g for 5 minutes and the supernatant was discarded. Following this the bacteria was suspended in 40mL of ice cold 60mM $CaCl_2$ and incubated on ice for 30 minutes. Bacteria was then centrifuged at 1100g for 5 minutes at 4°C and the pellet was then resuspended in 8mL of ice cold 60mM $CaCl_2$ and aliquoted into 100µL stocked in pre-chilled 1.5mL centrifuge tubes and frozen on dry ice before storage at -80°C. To verify and test efficiency new stock, a sample of new and old stock of bacteria were streaked onto ampicillin LB plates and incubated at 37°C overnight. In parallel 10ng of pUC19 control vector was transformed into 25µL of old and new stock of bacteria and incubated at 37°C overnight. No colony should be observed on ampicillin plates and the number of CFU in the transformation was compared.

3.3.3. Mini-prep of plasmids DNA

After successful transformation a single colony of bacteria was used to inoculate 4mL of LB broth containing appropriate antibiotic for selection, and incubated overnight at 37°C with shaking to create a starter colony. The following day 3mL of culture was pelleted in a 1.5mL centrifuge tube and resuspended in 100µL buffer P1 (Qiagen) followed by alkali lysis by the addition of 100µL of buffer P2 (Qiagen), inverted 6 times and incubated at room temperature for no longer than 5 minutes. 100µL Buffer P3 (Qiagen) was added to neutralise the solution and the tube inverted 6 times and incubated on ice for 5 minutes. The chilled solution was centrifuged at 16,000g for 10

minutes. The supernatant was transferred to a clean 1.5mL centrifuge containing 500µL 200 proof EtOH and place at -20°C for up to an hour. The DNA was pelleted by centrifugation at 16,000g for 10 minutes at room temperature. The EtOH removed and the pelleted washed with 750µL 70% EtOH and allowed for air-dry. The pellet was resuspended in 50µL of nuclease free dH_2O. The purified plasmid was subjected to restriction enzyme digestion and electrophoresis to ensure the correct plasmid was present and then stored at -20°C. The remaining starter culture was used to create a glycerol stock to store vectors long term.

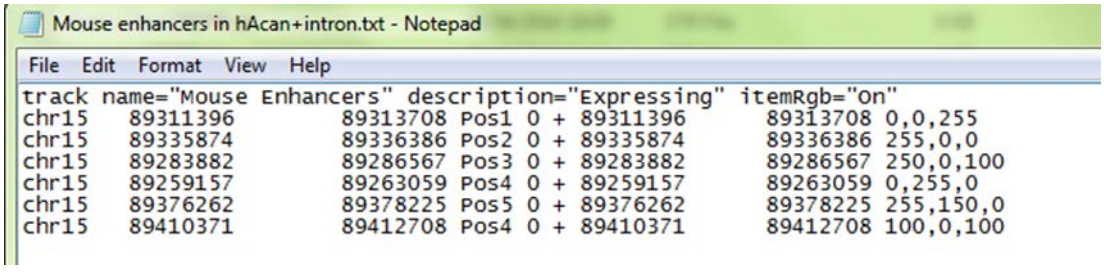
3.4. Identification of limb specific aggrecan enhancers

The Encyclopedia of DNA Elements project (ENCODE) provides publically available data sets on functional elements of the genome that can be viewed on their hub or at The University of California at Santa Cruz (UCSC) genome browser (https://genome.ucsc.edu/). These data set has advanced throughout this project into difference accessions and experiments which can be annotated and utilised for our gene of interest, additionally, custom tracks taken from datasets available from the Gene Expression Omnibus (GEO) can be edited and visualised on the browser, this has enabled us to focus our efforts of characterising *Acan* enhancers on areas of particular interests. Initially, the mouse assembly July 2007 mm9 (NCBI37/mm9) was used to identify sequences up to 100kb upstream of *Acan* of interest. Epigenetic signatures H3K4me1, H3K27ac, H3K4me3 in E14.5 mouse limbs detected by ChIP-Seq was selected from the ENCODE Ludwig Institute for Cancer Research (ENCODE/LICR) project track data, DNase I hypersensitivity in E11.5 mouse limb buds, from the ENCODE project University of Washington (UW DNaseI HS) and high conservation throughout vertebrate genomes were used as selection criteria's for the possibility of chondrocyte specific enhancers. At later stages of the work, to compare between developmental ages the mouse assembly December 2011 (GRCm38/mm10) was used as the reference assembly, to compare histone marks from ChIP-seq data at E12.5, E13.5 and E15.5 from Bing Ren's lab (UCSD) found on the ENCODE project experiment matrix database (https://www.encodeproject.org/matrix/?type=Experiment). To determine if other prediction software could identify enhancers which may have been

overlooked in our initial screen the mm10 mouse assembly and limb specific DNase1 HS and H3K27ac, using ENCODE annotations (http://zlab-annotations.umassmed.edu/). These prediction was compared to Sox9 data and the mm9 assembly. Furthermore, GEO data sets for Sox9 ChIP-seq and histone ChIP-seqs were downloaded as .Bed, .wig or .bedGraph files and added as custom tracks using UCSC genome browser custom track tab, dataset consisted of Sox9-ChIP from primary mouse rib chondrocytes and embryonic E17.7 nasal chondrocytes series: GSE69111 (Ohba et al., 2015), rat chondrosarcoma sox9-ChIP series: GSE70144 (Liu and Lefebvre, 2015), sox-9 ChIP from E12.5 limbs and heart valves accession GSE73225 (Garside et al., 2015), H3K122ac and H3K64ac for mouse embryonic stem cells and H3K122ac, H3K27ac and H3K4me1 for human bone marrow cell line K562 from dataset: GSE66023 (Pradeepa et al., 2016), K56ac from primary mouse embryonic stem cells series: GSE47387 (Tan et al., 2013). Additionally, to compare the zebrafish data which screened the *Acan* locus for enhancers from Hu *et* al 2012, were mapped to the GRCh37/hg19 human assembly and converted to the NCBI37/mm9 assembly and a custom track for the genome browser expressing and screened was created (Hu et al., 2012). Due to the strong conservation of the *Acan* enhancers we mapped each one to the human (GRCh37/hg19) accession to determine the chromatin state in different cell types (Ernst et al., 2011).

The genome browser tool LiftOver (https://genome-euro.ucsc.edu/cgi-bin/hgLiftOver?hgsid=213826534_FQkMJ6mDgKnn5Xme0ooE1AlOaNWi) was used to annotate datasets from different species or accessions. Data points were created, converted and then manually formatted into the Browser Extensible Data (.BED) format and each point colour coordinated using the RGB colour model.

Figure 16: General editing window for custom tracks



3.5. Standard manipulation of nucleic acids

3.5.1. Oligonucleotides

To design and verify the primers used in PCR based studies web based program primer blast was used (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). Additionally, primer temperatures prediction and primer dimer estimation were assessed using the Multiple Primer Analyzer (https://www.thermofisher.com/uk/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/multiple-primer-analyzer.html#/legacy=www.thermoscientificbio.com). Primers were obtained from Eurofins genomic and resuspended in nuclease-free water to a concentration of 100pmol/ µL.

3.5.2. Polymerase Chain Reaction (PCR)

All DNA fragments used for cloning were amplified by PCR from 100ng of mouse genomic DNA (Bioline, BIO-35027 and Promega, G3091). For Gateway™ based cloning a non- high-fidelity DNA *taq* polymerase, REDExtract™-N-Amp ReadyMix™ (Sigma, E3004) was used to utilise the ability of the *taq* polymerase to perform A' tailing to allow for TA cloning. Reaction parameters of 94°C for 3 minutes to denature the DNA, 30 seconds at the specific primer annealing temperature (58°C - 62°C), extension of 2 minutes at 72°C for 30 cycles a prolonged final extension time of 10 minutes and a soak step at 4°C. Q5® High-Fidelity DNA polymerase (NEB, M0491S) was used for cloning the *Acan* regions for luciferase based reporters. The reaction consisted of in a final volume of 25µL, 1X Q5 Reaction buffer, 1X Q5 High GC Enhancer, 200µM of dNTPs, 0.5µM of each primer (forward and reverse) and 0.02U of polymerase. The DNA was denatured at 98°C, primer annealed depending on specific annealing temperature (60°C- 68°C) and a final extension at 72°C. All PCR reactions was verified by agarose gel electrophoresis.

3.5.3. PCR clean up

Taq polymerase, $MgCl_2$, buffer and dNTPs were removed from PCR products by Monarch® PCR & DNA clean up kit (5µg) (NEB). For 20µL of PCR reaction 100µL of DNA Clean up binding buffer was added and mixed. Samples were placed into the spin column and centrifuged at 16,000g at room temperature for 1 minute and the flow-through was discarded. 200µL of DNA wash buffer was added and the column spun again at 16,000g for 1 minute and repeated. The column was placed into a clean 1.5mL centrifuge tube and 10µL of warmed (50°C) DNA elution buffer was added to each column and incubated at room temperature for 5 minutes then centrifuged at 16,000g for 1 minute to elute.

3.5.4. Restriction enzymes digestion

All restriction enzymes used were purchased for New England Biolabs (NEB). Restriction digestion was carried out in 20µL volume with 10 units of enzymes per 1µg of vector used following the manufacturers recommendations. For verification of inserts 1-hour digests were carried out, for cloning overnight digestions was used. Following digestion enzymes were heat inactivated when possible at either 65°C or 80°C for 20 minutes or gel purified.

3.5.5. DNA Dephosphorylation

After restriction digestion, vectors were desphosphorylated to prevent self-ligation. Shrimp alkaline phosphatase (rSAP) (NEB) was used, alternatively Antarctic phosphatase (NEB) replaced rSAP. rSAP was added directly to the restriction mix whereas Antarctic phosphatase buffer was used to supplement the restriction mix and the volume brought up to 30µL. The phosphatase was added to each reaction at 1-5 units per 1pmol of DNA ends based on the calculation:

$$\mu g\ DNA \times \frac{pmol}{600pg} \times \frac{10^6 pg}{1\mu g} \times \frac{1}{N} \times 2 \times \frac{kb}{1000bp} = pmol\ DNA\ ends$$

N is the nucleotides in kilobases, 660pg/mol is the average molecular weight of a nucleotide pair, 2 is the number of ends, Kb/1000bp is the conversion factor.

Reactions were then incubated at 37°C for 1hour for rSAP and 30minutes Antarctic phosphatase then heat inactivated at 65°C for 5 minutes for rSAP or 80°C for 2 minutes for Antarctic phosphatase.

3.5.6. Agarose gel electrophoresis

Gels were made as 0/8% (w/v), 1% (w/v), 1.5% (w/v) or 2% (w/v) using low electroendosmosis (EEO) agarose (Sigma) added to 1X Tris- acetate- EDTA (TAE) buffer and microwaved for 1 minute and 20 seconds for every 100mL. Ethidum bromide (EtBr) (0.2µg/µL) or SafeView nucleic acid stain (5µL) (NBS Biologicals) were used to allow for visualisation under ultraviolet (UV) exposure. Electrophoresis was carried out at a constant electrostatic field of 10V/ cm and gels were visualised on a UV transilluminator (BioRad ChemiDoc XRS) at 254nm for analysis or on a UV transilluminator platform at 312nm for cloning.

3.5.7. Purification of DNA from agarose gel

DNA were isolated from agarose gels using the QIAquick gel extraction kit (Qiagen, 28706) following manufacturers instruction and eluted with pre- warmed (50°C) nuclease-free water.

3.5.8. Cloning of *Acan* enhancers *in vivo* constructs

Enhancer sequences were generated by PCR using primers designed with flanking 100bp 5' and 3' identified from the ENCODE analysis. PCR products were gel purified on a 1% agarose gel. Fragments were then cloned into the pCR™8/GW/TOPO® vector using the pCR™8/GW/TOPO® TA Cloning Kit (Invitrogen, K2500-20) following manufacturer's instructions. In short, 2µL of DNA freshly extracted DNA product, 1µL salt solution, 0.5µL TOPO® vector was gently mixed with 2.5µL dH_2O, touch down centrifuged and incubated at 22-25°C for 30 minutes for short fragments of up to 1.5kb and 1 hour for >1.5kb. 4µL of the TOPO® reaction was transformed into TOP10 bacteria and incubated overnight at 37°C on streptomycin (50µg/mL) LB- agar plates. The following day starter cultures were created and then minipreped.

Name	Sequence 5’ – 3’
Acan +60kb F	AAT CAC CTG CAC AGA CCG TGA
Acan +60kb R	TGT AGC CTG TGC TTG TAG GTG T
Acan +55kb F	ATA AAC CAG GCG TGA TGC TAC ATA CCT GTA
Acan +55kb R	AAG GCA GGA GGC TGC AGG AAA GAT
Acan +26kb F	CTG AGT CTT GAC AGA GCT GCT A
Acan +26kb R	GAT TGG TCA GTA GTG GAG CAC A
Acan +26kb Small F	TCC ATC TCA GGC CTT CCT CT
Acan +26kb Small R	AGG GCA TAG AGG CAC ATG AC
Acan -10kb F	GGT GAC ACC CAA ACT AAC AGC
Acan -10kb R	ATC TCA CAC GTG GGC ATT TGC
Acan -35kb F	GGC TAG ACC AGT AGA ACC CA
Acan -35kb R	TGC ATC ATT ACA GTC TTT CTT CAG
Acan -35kb Small F	CGC ATG CAC ACC GCC CTC CT
Acan -35kb Small R	CTT CAG CAG TCA GGC AGG GGA
Acan -65kb F	ATG TTC AGA CAG CTC CAA CCC
Acan -65kb R	GGG TTG TGT TAA TAG GCA CGG
Acan -87kb F	CAG ACT GTA TTT CCC GAG TAC CT
Acan -87kb R	TGG AAG AGG GGC AAA ATG AGA
Acan -87kb Small F	GTG AGG ACT GAA GGG AAC GA
Acan -87kb Small R	CCT GGG ACA AGG TTT TGT GG
***Acan* -68kb F**	**CGT GAG TAT CCA GAG TCT CCT**
***Acan* -68kb R**	**GAA GAT GCT GGC ATT GTG CTT**
***Acan* -108kb F**	**GGG TGT GCC TTA TTG AAT GGC**
***Acan* -108kb R**	**AGA AGC ATC TTG AGT GAA GGG T**
***Acan* -115kb F**	**AGC CCC CTA GAA CCT TGC T**
***Acan* -115kb R**	**ACC TTT TCA AAT GGA AAT TCAA CGC**
***Acan* -128kb F**	**CCC TGA CCA ACG TCC CAT C**
***Acan* -128kb R**	**GCT GGG ACT GAT TCT GCA CTT**

In bold are the constructs that were created but never injected

Cloning success and correct orientation of insertion was checked and verified by restriction enzyme digestion and Sanger sequencing, sequencing was performed by SourceBioscience. The enhancer sequences were transferred from the pCR8/GW/enhancer entry vector to the HSP68/*LacZ*/Gateway® destination vector (Pennacchio et al., 2006) using Gateway LR Clonase II Enzyme mix (Life Technologies, 11791-020). The following reaction was set up in a 1.5mL centrifuge tube:

Reagent	Volume
HSP68/*LacZ*/Gateway® vector	37ng/µL
Entry vector	37ng/µL

LR Clonase II enzyme was thawed for 2 minutes on ice then vortexed for 2 seconds twice, 0.5µL of the enzyme was added to the reaction mix, mixed and then incubated at 25°C for 1 hour or up to 16 hours for larger fragments. The reaction was stopped by the addition of 0.25µL of Proteinase K solution (2µg/µL) and incubated at 37°C in a water bath for 15 minutes. The entire reaction was transformed in the TOP10 bacteria and plated onto ampicillin selection plates. The following day starter cultures and then miniprep was performed.

3.6. Generation transgenic mice

All animal procedures were approved by local ethical review and licensed by the UK Home Office and licensed under the Animal (Scientific Procedures) Act 1986, project licence number (PPL) 70/7288. Mice used for transgenic studies are mixed background C57BL/6 X CBA B6CBAF1. Animals were housed in the University of Liverpool Biomedical service unit (BSU) a specific pathogen free (SPF) mouse facility, in 12/12 hour light/dark cycle, 45 ± 10% humidity and a constant temperature of 22 ± 2°C.

3.6.1. Preparation of DNA for microinjection

In preparation of oocyte microinjection constructs 5µg of expression vector was digested overnight at 37°C with restriction enzymes XhoI and NotI or EagI. Fragment of interest were gel purified using 0.8% agarose with 0.5µg/mL EtBr and the Qiaquick Extraction kit. The concentration was determined and diluted to 2ng/µL in embryo water (Sigma) and filter purified using Durapore®-PVDF 0.22µM centrifuge filters (Merck Millipore LtD), and stored at -20°C. DNA was warmed to room temperature prior to injection. Primary responsibly of the mice resided with myself with the help of the staff at the BSU for daily care.

Transgenic mice were generated by oocyte microinjection detailed previously (Sacilotto et al., 2013) and carried out by Dr Ke Liu.

3.6.2. Superovulation of egg donors

Superovulation was performed on F1 females aged between 4-6 weeks old. 5 IU of follicle stimulating hormone was injected intraperitoneally, after 46hours 5 IU of luteinising hormone was administered via intraperitoneal injection. Females were then mated with F1 males overnight. The following day copulation plugs were checked to identify mated females. Females were anesthetised and oviducts were removed and dissected in M2 media (Millipore, Specialty Media, EmbryoMax®). Eggs were transferred to clean media via mouth pipetting. Cumulus cells were removed by hyaluronidase (300µg/mL) treatment in M2 media with gentle shaking until visibly detached from the egg surface. Eggs were then rinsed and transferred to M16 media (Millipore, Specialty Media, EmbryoMax®) ready for injection.

3.6.3. Injection of fertilised eggs and transfer to pseudopregnant recipients

Injections of fertilised eggs were carried out by Dr Ke Liu (University of Liverpool) and has been previously described (Hogan et al., 1994), with video aid found in the Journal of Visualized Experiments (JoVE) (Stein and Schindler, 2011). Injection pipettes piece the

oocyte avoiding the nucleus and DNA was injected. Undamaged eggs were transferred to clean M16 media and incubated until transfer into pseudopregnant F1 females.

Pseudopregnant females were generated by mating vasectomised F1 males overnight the day before injection of fertilised eggs. Copulation plugs were checked and females were used 1 day post-coitum. Females were anaesthetised with 50µL hypnorm and 50µL hypnoval by intraperitoneal injection. Eggs were microinjected at the 1 cell stage and transferred through the infundibulum to the oviducts. On average 20-30 eggs were transferred per female.

3.7. Transgenic mouse analysis

3.7.1. X-gal staining of mouse embryos E15.5 and adult tissue

3.7.1.1. Solutions

1M Sodium Phosphate (NaPi) Buffer pH7.4

138g of sodium dihydrogen phosphate ($NaH_2PO_4 – H_2O$) (*Sigma, 71507)* was dissolved in 1 litre of ddH_2O and pH 7.4. 142g sodium phosphate dibasic (Na_2HPO_4) (*Sigma, S7907*) was dissolved in 1 litre of ddH_2O, pH 7.4. To make 1M NaPi buffer 423mL of $NaH_2PO_4 – H_2O$ and 577mL of Na_2HPO_4 was mixed together.

3.7.1.2. X-gal Fixative

Reagent	For 300mL
0.2% Glutaraldehyde	2.4mL of 25% solution (*Sigma, G6257)*
0.1M Na Pi buffer pH7.3	30mL of 1M stock
5mM ethylene glycol tetraacetic acid (EGTA), pH 8.0	15mL of 0.1M stock (*Sigma, E0296*)
2mM $MgCl_2$	600µL of 1M stock (*Sigma, M1028*)
2% formalin	60mL of 10% stock or 15ml of 40% stock
ddH_2O	Up to 300mL

3.7.1.3. Rinse solution

Reagent	For 1 Litre
0.1M NA Pi buffer pH 7.3	100mL 1M stock
2mM $MgCl_2$	2mL of 1M stock
0.1% Na deoxycholate	1g (*Sigma, 30970)*
0.2% NP40 substitute	2mL (*Sigma, 74385)*
ddH_2O	Up to 1 litre

3.7.1.4. Stain Solution

Reagent	For 100mL
Rinse solution	100mL
Potassium ferricyanide ($K_3Fe(CN)_6$) (Sigma, 702587)	165mg
Potassium ferrocyanide ($K_4Fe(CN)_6$) (Sigma, 60279)	184mg

Stored in the dark at 4°C

3.7.1.5. X-gal stock (25mg/mL)

0.5g of 5-bromo-4-chloro-3-indolyl-β-d-galactopyranoside *(VWR, 437132J)* (X-gal) was suspended in 20mL of *N, N*- dimethylformamide (*Sigma, D4551*) and store at -20°C.

3.7.1.6. X-gal staining

Transgenic mouse embryos were collected at E13.5 and E15.5. After dissection and removal of yolk sac and placenta, samples were rinsed in ice-cold 1X PBS and fixed in 6X the volume of ice cold X-gal fix for 45 minutes whilst rotating. After fixation embryos were washed three times in rinse solution, embryos were then stained overnight in the dark at room temperature in 1 mg/ml X-gal solution containing 5 mM $K_4Fe(CN)_6$, 5 mM $K_3Fe(CN)_6$, 0.1% sodium deoxycholate, 0.2% Nonidet P-40, 2 mM $MgCl_2$ and 0.1 NaPi buffer. After staining, embryos were rinsed through a series of three 30 minutes washes, and then fixed for 24 hours in 6X the volume of the embryos in 10% neutral buffered formalin (NBF) at 4 °C.

Eight week old transgenic mice were culled by cervical spinal dislocation and dissected. All bones and cartilaginous tissue were taken, muscle cleaned off and then fixed in 6X the volume of x-gal fix for 2 hours at room temperature whilst rotating. Following 3X 30 minutes washes in rinse solution tissue was stained in x-gal stain solution overnight in the dark. The following day, samples were washed, imaged and fixed in 10% NBF for 24 hours before decalcification with 10% EDTA for 2 weeks.

Imaging of whole embryos and adult tissue was performed using the Olympus SZX12 equipped with the Qimaging QIClick™ CCD camera and QCapture Pro software. For each enhancer, embryos and adult tissue were also sectioned for histological analysis to investigate localisation of X-gal staining. Samples were process by dehydration with ethanol washes, cleared by xylene and paraffin wax-embedded following table

3.7.1.7. Clearing and imaging of whole mount embryos

To visualise whole mount x-gal stained embryos with better contrast and visualise the skeletal elements better, embryos up to E15.5 were fixed in 6X the volume of 10% NBF overnight at 4°C and then 70% EtOH at 4°C overnight. Embryos were then cleared in 15 mL of 20%, 50%, 80% and 100% glycerol in 1% potassium hydroxide (KOH) solution for 3 days at room temperature with agitation.

Images were taken as previous on the Olympus SZ12 in bright field with embryos suspended in 100% glycerol.

3.7.2. Tissue processing

Processing of LacZ stained Embryo up to E14.5:

Reagent	Immersion Period (Minutes) under vacuum
70% Ethanol	30
90% Ethanol	30
100% Ethanol	10
100% Ethanol	10
100% Ethanol	10
100% Ethanol	10
Xylene	10
Xylene	10
Wax	100
Wax	100

For larger LacZ stained embryos E14.5 to E18.5:

Reagent	Immersion Period (Minutes) under vacuum
70% Ethanol	30
90% Ethanol	30
100% Ethanol	10
100% Ethanol	10
100% Ethanol	10
100% Ethanol	10
Xylene	30
Xylene	30
Wax	120 (2hours)
Wax	Overnight

For decalcified bone:

Reagent	Immersion Period (Minutes) under vacuum
70% Ethanol	15
95% Ethanol	60
100% Ethanol	60
100% Ethanol	60
100% Ethanol	60
100% Ethanol	60
Xylene	60
Xylene	60
Wax	120
Wax	Overnight

6-µm sections were prepared using the Micron HM355S microtome with cool-cut and tissue transfer system (Thermo Scientific), de-waxed, and counterstained with Eosin (Leica). Histo-Clear (National diagnostics) was used in place of xylene to prevent the leeching of the X-gal staining. Eosin counter-stained sections were imaged on the Olympus BX60 using the CellD software. Images were collated and processed using Corel Paintshop Pro X6.

3.7.3. BR1 and HAPLN3 antibody staining for aggrecan and link protein 3

The antibody, BR1, was generated by Professor Tim Hardingham and donated by Alan Murdoch, is a polyclonal rabbit anti-pig antibody raised against the G1 domain of aggrecan. It was used to determine the endogenous expression pattern of ACAN in embryonic and 8 week old mouse tissues. The HAPLN3 antibody HPA039237 from Sigma, is a polyclonal anti-human raised in rabbit and was used to determine expression of link protein 3 in 8 week old mouse tissue. The Dako EnVision+ System-HRP (DAB) for use with rabbit primary antibody was used to visualise the BR1and HAPLN3 antibody.

Tissue sections were dehydrated overnight at 37°C, then dewaxed in 2X xylene for 5 minutes each. The samples were rehydrated in 2X 100% EtOH, 90% EtOH and 75% EtOH for 5 minutes each, rinsed in distilled water for 30 seconds and washed in 3x PBS for 5 minutes each with agitation. Slides were then incubated with bovine testicular hyalouronidase (1000U/mL made up in PBS) (Sigma) for 1 hour at 37°C. Slides were then rinsed in 3x PBS for 5 minutes each with agitation then incubated with peroxidase block for 10 minutes and washed again in 3x PBS for 5 minutes each. 1% non-fat milk made up in PBS was then added to the tissue for 1 hour at room temperature. Slides were then incubated with BR1 or HAPLN3 antibody (1:1000 or 1:500 made up in Dako antibody diluent) at 4°C for 16 hours. Slides were then rinsed in 3x PBS for 5 minutes then incubated with the secondary antibody (peroxidase labelled polymer conjugated to goat anti-rabbit immunoglobulins) and then washed again in 3x PBS for 5 minutes each. Peroxidase substrate (DAB+ chromogen made up in DAB+ substrate buffer) was added for

30 seconds up to 5 minutes. Slides were then placed in running tap water for 10 minutes. Counterstained with Harris haematoxylin for 3 minutes and rinsed in running tap water for 10 minutes then dehydrated in a series of 70% EtOH, 90% EtOH, 2X 100% EtOH for 2 minutes each then 2x xylene for 5 minutes each and coverslipped using PERTEX® mounting media (Histolab).

3.7.4. Genotyping of transgenic mice

3.7.4.1. Solutions

3.8.2.1.1 Lysis buffer

Reagents	Volume
50mM Tris-HCL (pH8)	4.5mL of 2M stock
0.1M NacL	7mL of 3M Stock
1% SDS	8mL of 10%
20mM EDTA	7.2mL of 500mM Stock
dH_2O	144.3mL

Autoclaved and store at room temperature

3.8.2.1.2 Proteinase K Stock

10mg/mL in 50% Glycerol stored at -20°C

3.7.4.2. Proteinase K based extraction of DNA

Transgenic lines were created for each of the expressing *cis*-acting elements. Ear notches of adult mice and tail samples from embryos were used for genotyping. Samples were digested in 100µL lysis buffer with proteinase K (10mg/µl) at 55°C overnight. To precipitate the proteins out of solution 100µL 3M ammonium acetate ($C_2H_3O_2NH_4$) was added to the lysed tissue and vortexed briefly. The solution was incubated on ice for 5 minutes prior to centrifugation at 16000g for 10 minutes at room temperature. 100%

chloroform was added to allow for phase separation and centrifuged at 16,000g for 10 minutes at room temperature. The upper phase was taken and placed into a fresh centrifuge tube containing 500µL of ice cold 200 proof 100% EtOH, the tubes were inverted twice and incubated at -20°C for 30 minutes to an hour to precipitate the DNA. The tubes were then inverted twice and centrifuged for 10 minutes at room temperature, the 100% EtOH was removed and replaced with 750µL 70% EtOH to wash the DNA pellet, the tubes were briefly centrifuged at 16,000g. The EtOH was then removed and the DNA pellet allowed to air-dry for up to 2 hours. The pellet was resuspended in TE buffer overnight at 4°C before long term storage at -20°C.

3.7.4.3. Genotyping of Transgenic mice using PCR

50ng/µL of purified DNA was used for multiplex PCR using the GoTaq®G2 Flexi DNA polymerase (Promega, M7805) supplemented with 2.5mM $MgCl_2$ using *LacZ* specific primers with internal control and verified using a second *LacZ* primer set.

Primer Name	Sequence 5' – 3'	Species	Annealing Tm	Target	Product Size
HELp008	TGG ACA GGA CTG GAC CTC TGC TTT CCT AGA	Mouse	68	*Fabp2* (Internal)	
HELp009	TAG AGC TTT GCC ACA TCA CAG GTC ATT CAG	Mouse	68	*Fabp2* (Internal)	194bp
HELp012	GTT GCA GTG CAC GGC AGA TAC ACT TGC TGA	Mouse	68	*LacZ*	
HELp013	GCC ACT GGT GTG GGC CAT AAT TCA ATT CGC	Mouse	68	*LacZ*	389bp
prIL010	GAA CTA CCG CAG CCG GAG A	Mouse	61	*LacZ*	
prIL011	GGT AAT GGT AGC GAC CGG C	Mouse	61	*LacZ*	1000bp

3.8. Characterisation of Acan enhancers

3.8.1. Identification of Transcription factor binding sites

Each sequence was aligned to other species conserved sequences using ClustalW and subjected to analysis using the Transfac software (BioBase) (Matys et al., 2006) and Transcription Factor Affinity Prediction (TRAP) webtools (http://trap.molgen.mpg.de/cgi-bin/home.cgi) (Thomas-Chollier et al., 2011) and conservation for the most represented sites to take forward for analysis. Sequences of around 18-30 nucleotides long were used for EMSA experiments to determine TFs binding affinity.

3.8.2. Electrophoretic mobility shift assay

3.8.2.1. Solutions

Black's buffer (10X): 400mM KCl, 150mM, HEPES (pH7.9), 10mM EDTA, 5mM DTT, 50% glycerol

2X anneal buffer: 100mM Tris pH7.6, 10mM DTT, 20mM $MgCl_2$, 2mM spermidine

Elution buffer: 40mM KCl

10X Tris-Boric-EDTA (TBE) buffer: 890mM Tris-borate, 890mM boric acid, 20mM EDTA

3.8.2.2. Oligonucleotides for EMSA

Probes and double stranded oligos used in EMSA were designed to have a "CTAG" overhang on the 5' end of the forward and 3' of the reverse to allow for radiolabelling, the exceptions are the Rbpj-k control oligos that have "GG" instead of the "CTAG". Oligos were synthesised by Integrated DNA Technologies (USA) and resuspended in nuclease-free water to a concentration of 1mg/mL.

Name	Sequence 5'- 3'
Sox9 control F	CTAGGATCCAAAGCCCCATTCATGAGATCTG
Sox9 control R	CTAGCAGATCTCATGAATGGGGCTTTGGATC
Rbpj-k control F	GGGGTGAGGTCTATTTCCCACGACATACTTCC
Rbpj-k control R	GGGGGAAGTATGTCGTGGGAAATAGACCTCAC
Nfat control F	CTAGTGCTCTGAAACATTTTCCTCCTTCACAGT
Nfat control R	CTAGACTGTGAAGGAGGAAAATGTTTCAGAGCA
Acan -35kb Sox 1F	CTAGCTTGAACACTGCCTGCTGTCCCAACC
Acan -35kb Sox 1R	CTAGGGTTGGGACAGCAGGCAGTGTTCAAG
Acan -35kb Sox 1mut F	CTAGCTTCCTGACTGCCTGCACCCCCAACC
Acan -35kb Sox 1mutR	CTAGGGTTGGGGGTGCAGGCAGTCAGGAAG
Acan -35kb Sox 2F	CTAGAAAGGAACACAACCCTCCCTTGTGCCCTCCT
Acan -35kb Sox 2R	CTAGAGGAGGGCACAAGGGAGGGTTGTGTTCCTTT
Acan -35kb Sox 2mutF	CTAGAAAAGGGGCTCCGCCCTCCAGGGAGCCCTCCT
Acan -35kb Sox 2mutR	CTAGAGGAGGGCTCCCTGGAGGGCGGAGCCCCTTTT
Acan -35kb Sox 3F	CTAGGTGCCCTCCTGATATTTCCAGTAAAAAGTACCTT
Acan -35kb Sox 3R	CTAGAAGGTACTTTTTACTGGAAATATCAGGAGGGCAC
Acan -35kb Sox 3mutF	CTAGGTGCCCTCCTTGGGCCTCCCGTCCACGTACCTT
Acan -35kb Sox 3mutR	CTAGAAGGTACGTGGACGGGAGGCCCAAGGAGGGCAC
Acan -35kb Sox 4F	CTAGCCTTGAACACAATCTTGCTGCCACAGCCCTAA
Acan -35kb Sox 4R	CTAGTTAGGGCTGTGGCAGCAAGATTGTGTTCAAGG
Acan -35kb Sox 4mutF	CTAGCCTTGAGCGCGATCTTGCCGCCACAGCCCTAA

Acan -35kb Sox 4mutR	CTAGTTAGGGCTGTGGCGGCAAGATCGCGCTCAAGG
Acan -35kb Sox 5F	CTAGCCCTAATTATGTGTGAAATCATTTTTTTTAAAGC
Acan -35kb Sox 5R	CTAGGCTTTAAAAAAAATGATTTCACACATAATTAGGG
Acan -35kb Sox 5mutF	CTAGCCCTAATTATGTGGGCCCTCATTTTTTTTAAAGC
Acan -35kb Sox 5mutR	CTAGGCTTTAAAAAAAATGAGGGCCCACATAATTAGGG
Acan -35kb Sox 6F	CTAGAAAGCTTTAAACCAAAGCAACAAGCCCTA
Acan -35kb Sox 6R	CTAGTAGGGCTTGTTGCTTTGGTTTAAAGCTTT
Acan -35kb Sox 6mutF	CTAGAAAGCTCTGGCCCAACGCACACAGGCCCAT
Acan -35kb Sox 6mutR	CTAGATGGGCCTGTGTGCGTTGGGCCAGAGCTTT
Acan -35kb Sox 7F	CTAGCCCATACAGAGAGGCCCATTCTCCCAGCCC
Acan -35kb Sox 7R	CTAGGGGCTGGGAGAATGGGCCTCTCTGTATGGG
Acan -35kb Sox 7mutF	CTAGGGCTGGGAGCCTGGGCCCCTCCGTATGGG
Acan -35kb Sox 7mutR	CTAGGGCTGGAGCCTGGGCCCCTCCGTATGGG
Acan -35kb Sox 8F	CTAGGGTTACCTCAGGCTACACAGAGGCC
Acan -35kb Sox 8R	CTAGGGCCTCTGTGTAGCCTGAGGTAACC
Acan -35kb Sox 8mutF	CTAGGGTTACCTCCTCCTACTTGGAGCGCC
Acan -35kb Sox 8mutR	CTAGGGCGCTCCAAGTAGGAGGAGGTAACC
Acan -35kb Rbpj-A F	CTAGACACACCCTCCCTTGTGCCCTCCT
Acan -35kb Rbpj-A R	CTAGTCAGGAGGGCACAAGGGAGGGTGTGT
Acan -35kb Rbpj-A mutF	CTAGACACAACCCTCGGTTCGGCCTCCTGA

Acan -35kb Rbpj-A mutR	CTAGTCAGGAGGCCGAACCGAGGGTTGTGT
Acan -35kb Rbpj-B F	CTAGACCTTGAACACAATCTTGCTGCC
Acan -35kb Rbpj-B R	CTAGGGCAGCAAGATTGTGTTCAAGGT
Acan -35kb Rbpj-B mutF	CTAGACCTTGAACGTGGCTTGCTGCC
Acan -35kb Rbpj-B mutR	CTAGGGCAGCAAGCCAACGTTCAAGGT
Acan -35kb Rbpj-C F	CTAGCCTAATTATGTGTGAAATCATTTTTT
Acan -35kb Rbpj-C R	CTAGAAAAAATGATTTCACACATAATTAGG
Acan -35kb Rbpj-C mutF	CTAGCCTAATTATTTGTTAAATCATTTTTT
Acan -35kb Nfat1 F	CTAGCTAGCTGGAAAAGGAACACAA
Acan -35kb Nfat1 R	CTAGTTGTGTTCCTTTTCCAGCTAG
Acan -35kb Nfat1 mF	CTAGCTAGCTCAGTGGCCTTCAA
Acan -35kb Nfat1 mR	CTAGTTGAAGGCCACTGAGCTAG
Acan -35kb Rbpj-C mutR	CTAGAAAAAATGATTTAACAAATAATTAGG
Acan -87kb Sox 1F	CTAGGGAACGACAAAATGTCCT
Acan -87kb Sox 1R	CTAGAGGACATTTTGTCGTTCC
Acan -87kb Sox 1mF	CTAGGGAACGTCGGAATGTCCT
Acan -87kb Sox 1mR	CTAGAGGACATTCCGACGTTCC
Acan -87kb Sox 1_2F	CTAGGGAACGACAAAATGTCCTTACTGTCCCCAC
Acan -87kb Sox 1_2R	CTAGGTGGGGACAGTAAGGACATTTTGTCGTTCC
Acan -87kb Sox 1_2mF	CTAGGGAACGACATTATGTCCTTACTGGACCCAC

Acan -87kb Sox 1_2mR	CTAGGTGGGTCCAGTAAGGACATAATGTCGTTCC
Acan -87kb Sox 2_3F	CTAGGTCCTTACTGTCCCCACTTGTGGTCCA
Acan -87kb Sox 2_3R	CTAGTGGACCACAAGTGGGGACAGTAAGGAC
Acan -87kb Sox 2_3mF	CTAGGTCCTTCCGGTCCCCACTGGTGGTCCA
Acan -87kb Sox 2_3mR	CTAGTGGACCACCAGTGGGGACCGGAAGGAC
Acan -87kb Sox3F	CTAGCCCCACTTGTGGTCCAATCTGCCAAC
Acan -87kb Sox3R	CTAGGTTGGCAGATTGGACCACAAGTGGGG
Acan -87kb Sox3mF	CTAGCCCCACTGGGGGTCCATTCTGCCAAC
Acan -87kb Sox3mR	CTAGGTTGGCAGAATGGACCCCCAGTGGGG
Acan -87kb Sox4F	CTAGGAGAGTACATTTCCACGAGGC
Acan -87kb Sox4R	CTAGGCCTCGTGGAAATGTACTCTC
Acan -87kb Sox4mF	CTAGGAGAGTTCCTTTCCACGAGGC
Acan -87kb Sox4mR	CTAGGCCTCGTGGAAAGGAACTCTC
Acan -87kb Sox5F	CTAGTCCCAAACAACAGTACGGGC
Acan -87kb Sox5R	CTAGGCCCGTACTGTTGTTTGGGA
Acan -87kb Sox5mF	CTAGTCCCAAATTAGGGAACGGGC
Acan -87kb Sox5mR	CTAGGCCCGTTCCCTAATTTGGGA
Acan -87kb Sox6F	CTAGCATATTTGTCTTGGCC
Acan -87kb Sox6R	CTAGGGCCAAGACAAATATG
Acan -87kb Sox6mF	CTAGCATCGGGTCTTGGCC
Acan -87kb Sox6mR	CTAGGGCCAAGACCCGATG
Acan -87kb Sox7 Nfat3 F	CTAGAAAGACAGATTTTCCTTTCCC
Acan -87kb Sox7 Nfat3 R	CTAGGGGAAAGGAAAATCTGTCTTT

Acan -87kb Sox7 Nfat3 mF	CTAGAAAGACAGACGATCCACCCCC
Acan -87kb Sox7 Nfat3 mR	CTAGGGGGGTGGATCGTCTGTCTTT
Acan -87kb Sox8F	CTAGCCAGCCATTGGCCTCCGC
Acan -87kb Sox8R	CTAGGCGGAGGCCAATGGCTGG
Acan -87kb Sox8mF	CTAGCCAGCCCGGTTCCTCCGC
Acan -87kb Sox8mR	CTAGGCGGAGGAACCGGGCTGG
Acan -87kb Sox9F	CTAGGTGAGGCTTGTTTGGGATGG
Acan -87kb Sox9R	CTAGCCATCCCAAACAAGCCTCAC
Acan -87kb Sox9mF	CTAGGTGAGGACGGAACGGGATGG
Acan -87kb Sox9mR	CTAGCCATCCCGTTCCGTCCTCAC
Acan -87kb Sox10F	CTAGGATGGGCCACAAAACCTTGTCCC
Acan -87kb Sox10R	CTAGGGGACAAGGTTTTGTGGCCCATC
Acan -87kb Sox10mF	CTAGGATGGGCAACTTAACATCGTCCC
Acan -87kb Sox10mR	CTAGGGGACGATGTTAAGTTGCCCATC
Acan -87kb Nfat1 F	CTAGTGAGAGTACATTTCCACGAGGCCCAGG
Acan -87kb Nfat1 R	CTAGCCTGGGCCTCGTGGAAATGTACTCTCA
Acan -87kb Nfat1 mF	CTAGTGAGAGTACATCCACGAGGCCCAGTTT
Acan -87kb Nfat1 mR	CTAGAAACTGGGCCTCGTGGATGTACTCTCA
Acan -87kb Nfat2 F	CTAGATCTGGAATCTTCCAGCCCT
Acan -87kn Nfat2 R	CTAGAGGGCTGGAAGATTCCAGAT
Acan -87kb Nfat2 mF	CTAGATCTGGAATCCCCCAGCCCT

Acan -87kb Nfat2 mR	CTAGAGGGCTGGGGGATTCCAGAT
Acan -65kb Sox1F	CTAGAGAATGCGTCTTTGCTCGACCA
Acan -65kb Sox1R	CTAGTGGTCGAGCAAAGACGCATTCT
Acan -65kb Sox1mF	CTAGAGAATGCGTTGGTGCTCGACCA
Acan -65kb Sox1mR	CTAGTGGTCGAGCACCAACGCATTCT
Acan -65kb Sox2F	CTAGCAGGGCAGCATCAGAGGAGCGAGGA
Acan -65kb Sox2R	CTAGTCCTCGCTCCTCTGATGCTGCCCTG
Acan -65kb Sox2mF	CTAGCAGGGCAGTATCAGATTAGCGAGGA
Acan -65kb Sox2mR	CTAGTCCTCGCTAATCTGATACTGCCCTG
Acan -65kb Sox3F	CTAGGAGCTGGACAAGGGCAGGCTTTGTGCGCAGCG
Acan -65kb Sox3R	CTAGCGCTGCGCACAAAGCCTGCCCTTGTCCAGCTC
Acan -65kb Sox3mF	CTAGGAGCTGTACCAGGGCAGGCGGTGTGCGCAGCG
Acan -65kb Sox3mR	CTAGCGCTGCGCACACCGCCTGCCCTGGTACAGCTC
Acan -65kb Sox4F	CTAGTGTCTACTCTGAAGGAAAAT
Acan -65kb Sox4R	CTAGTGTCTACTCTGAAGGAAAAT
Acan -65kb Sox4mF	CTAGTGTAGACTCTGCCGGAGGAT
Acan -65kb Sox4mR	CTAGATCCTCCGGCAGAGTCTACA
Acan -65kb Sox5F	CTAGAGTCATTTCAAGGAATTGAATTTCAAGGAATCCT
Acan -65kb Sox5R	CTAGAGGATTCCTTGAAATTCAATTCCTTGAAATGACT
Acan -65kb Sox5mF	CTAGAGTCATTTCTTAAAACTGAAGCTCGGGGAATCCT
Acan -65kb Sox5mR	CTAGAGGATTCCCCGAGCTTCAGTTTTAAGAAATGACT
Acan +26kb Sox1Rbpj1 F	CTAGGCTTTCCTTCCCATTGAGAAAGGTCTCT
Acan +26kb Sox1Rbpj1 R	CTAGAGAGACCTTTCTCAATGGGAAGGAAAGC
Acan +26kb Sox1 Rbpj1 mF	CTAGGCTTTCCTTCCCAGGGAGCTAATTCTCT

Acan +26kb Sox1 Rbpj1 mR	CTAGAGAGAATTAGCTCCCTGGGAAGGAAAGC
Acan +26kb Sox2F	CTAGGAAAGGTCTCTCTGTCCACGAG
Acan +26kb Sox2R	CTAGCTCGTGGACAGAGAGACCTTTC
Acan +26kb Sox2mF	CTAGGAAAGGTTTATCTGGCTACGAG
Acan +26kb Sox2mR	CTAGCTCGTAGCCAGATAAACCTTTC
Acan +26kb Sox3F	CTAGTCCCTATTTTTCTTTACAAC
Acan +26kb Sox3R	CTAGGTTGTAAAGAAAAATAGGGA
Acan +26kb Sox3mF	CTAGTCCCTATTTCGACATACAAC
Acan +26kb Sox3mR	CTAGGTTGTATGTCGAAATAGGGA
Acan +26kb Sox4F	CTAGCTGAGAAAACACACACAAACACAAGAGC
Acan +26kb Sox4R	CTAGGCTCTTGTGTTTGTGTGTGTTTTCTCAG
Acan +26kb Sox4mF	CTAGCTGAGAATCCTCACACGCACACAAGAGC
Acan +26kb Sox4F	CTAGGCTCTTGTGTGCGTGTGAGGATTCTCAG
Acan +26kb Sox5F	CTAGCAGGGAGAAAAACAGAGCCAGCCCAAA
Acan +26kb Sox5R	CTAGTTTGGGCTGGCTCTGTTTTTCTCCCTG
Acan +26kb Sox5mF	CTAGCAGGGAGAATATCAGAGCCAGCGCAAA
Acan +26kb Sox5mR	CTAGTTTGCGCTGGCTCTGATATTCTCCCTG
Acan +26kb Sox6F	CTAGCAAACCCCACAGAGCACCATTGCATC
Acan +26kb Sox6R	CTAGGATGCAATGGTGCTCTGTGGGGTTTG
Acan +26kb Sox6mF	CTAGCAAACCTGACAGGCTACCATTGCATC
Acan +26kb Sox6mR	CTAGGATGCAATGGTAGCCTGTCAGGTTTG
Acan +26kb Sox7 Rbpj3 F	CTAGCAGCTTGTTCAGAGAGCGAGGGAAGAAACA
Acan +26kb Sox7 Rbpj3 R	CTAGTGTTTCTTCCCTCGCTCTCTGAACAAGCTG
Acan +26kb Sox7 Rbpj3 mF	CTAGCAGCTTGACCAGAGAGCGACGGTAGAAACA

Acan +26kb Sox7 Rbpj3 mR	CTAGTGTTTCTACCGTCGCTCTCTGGTCAAGCTG
Acan +26kb Rbpj2 F	CTAGTTTCATTCCTTCCCATTTCCTGGTT
Acan +26kb Rbpj2 R	CTAGAACCAGGAAATGGGAAGGAATGAAA
Acan +26kb Rbpj2 mF	CTAGTTTCATTCCTTCAAATCGTCCTGGTT
Acan +26kb Rbpj2 mR	CTAGAACCAGGACGATTTGAAGGAATGAAA

3.8.2.3. Making protein for EMSA

Proteins for indirect competition EMSAs were made using the TNT® T7/SP6 coupled reticulocyte lysate system (Promega, L5020) for Sox9 using 4XFLAG-tagged SOX9 expression plasmid (Lefebvre et al., 1998), Nuclear factor of activated T-cells (NFAT) core binding domain from pCITE 2B vector, recombination signal binding protein for Ig kappa J (Rbpj-κ) from pcDNA3.1 plasmid using T7 following manufacturer's instructions. In short, 50µL reaction volume was assembled (shown in table below) and incubated at 30°C for 1 hour and 30 minutes. Unprogrammed protein was used as controls to non- specific binding was generated by replacing expression vector with an empty.

Reagent	Volume
TNT® Rabbit Reticulocyte lysate	25µL
TNT® reaction buffer	2µL
TNT® T7 RNA polymerase	1µL
Amino acid mix minus Leucine, 1mM	0.5µL
Amino acid mix minus methionine, 1mM	0.5µL
Plasmid DNA	1µg
ddH_2O	Up to 50µL

3.8.2.4. Generating radioactively labelled probes for EMSA

Double stranded oligonucleotides were labelled with [α-^{32}P]-dCTP (3000Ci/mmol 10mCiml, 250µCi, EasyTide, Perkinelmer) using DNA polymerase I Large (Klenow) (Promega) to fill in overhanging ends and purified on non-denaturing acrylamide gel (40% acrylamide/bis 19:1, Biorad).

3.8.2.5. Preforming the shift assay

A competitive shift assay consisted of 20µL binding reactions with 2µg protein, 0.48µg of poly (dG-dC) for Sox9 and poly (dI-dC) for RBPJ- κ and 2µL 10X binding buffer. 100-fold excess or increasing concentration of competitor oligonucleotides was added to the binding reaction in a volume of 1µLand incubated at 30°C for 20 minutes. Radiolabelled probes were added at an activity of 40,000cpm/µL and incubated for a further 40 minutes. Gels were electrophoresed on a 6% non-denaturing polyacrylamide gel at 4°C for 2 hours. Gels were dried and exposed to X-ray films for 12 to 72 hours and then processed.

3.8.3. Generation of mutated binding sites

Mutant forms of the binding sites that were identified were generated by identifying the core binding sequence using MotifMap (http://motifmap.ics.uci.edu/) and single nucleotides were replaced to diminish the binding. The mutant forms were examined in EMSAs and then taken for *in vivo* studies.

Mutant versions of the -35kb and -87kb enhancer were generated as custom-made, double-stranded linear DNA fragments synthesised by GeneArt® Strings™ (Life Technologies) and were treated as PCR products used for cloning and generation of transgenic mice. The -87kb mutant TF binding constructs were never analysed in this study *in vivo* although they were generated.

3.8.4. *In vitro* analysis of Acan enhancers

To unravel the role multiple enhancers play in regulating the response of *Acan* transcription we created luciferase based reporters to evaluate in cell lines.

3.8.5. Generation luciferase reporter constructs

The +26kb, -10kb, -35kb, -65kb and the -87kb regions enhancer elements were examined *in vitro* by expression vectors containing luciferase. Enhancer regions were amplified from genomic DNA using Q5 High Fidelity polymerase (NEB) to generate blunted PCR products. PCR products were cleaned up using the Monarch® PCR & DNA clean up kit (NEB) prior phosphatase treatment These were then treated using T4 polynucleotide kinase (T4 PNK) (NEB) in the following reaction in 0.2mL PCR reaction tubes:

Reagent	Volume
PCR product	Up to 1µg
10X T4 ligase	2µL
T4 PNK	1µL
dH_2O	Up to 20µL

The reaction was incubated at 37°C for 1 hour then heat inactivated at 65°C for 20 minutes in a thermal cycler. Phosphatased inserts were blunt end cloned into the pGl3-promoter vector that has been digested with *SmaI* and dephosphorylated with rSAP or Antarctic phosphatase. *Acan* B small pGl3 – promoter was generated by extraction of the regions by restriction digestion of the *Acan*_B_wt pMa that was created by Life technologies GeneArt plasmid generation using *KpnI* and *SacI* overnight and gel purified. The region was then inserted into the pGl3- promoter that had been digested with *KpnI* and *SacI* overnight and dephosphorylated with rSAP.

Ligations were carried out at a 4:1 molar ratio of insert to vector. To estimate the concentration of construct, 2µL of sample was ran on a 1% agarose gel alongside 0.5µg and 1µg of lambda DNA-*HindIII* digest ladder and visualised, samples were also nanodropped for comparison. Amount of insert in a 10µL ligation mix was determined based on the calculation:

$$mass\ of\ insert\ (g) = desired\ \frac{insert}{vector}\ molar\ ratio\ \times mass\ of\ vector\ (g) \times ratio\ of\ insert\ to\ vector\ lengths$$

Ligations were carried out in the following reaction:

Reagent	Volume
10X T4 Ligase Buffer	1µL
Vector	1µL
Insert	X
T4 DNA Ligase (NEB)	1µL
For blunt end SmaI	1µL
ddH_2O	Up to 10µL

Blunt end ligations was carried out in the presence of the blunt end enzyme to prevent vector re-ligation. Ligations were incubated at 16°C overnight then heat inactivated at 65°C for 10 minutes. 5µL of ligation mix was used to transform into competent bacteria and plated onto ampicillin selective plates. After miniprep vectors were checked for correct insertion using restriction enzyme digestion.

3.9. Cell Culture

Two chondrocytic cell lines were used to study the effects each *Acan* enhancer has *in vitro*. All cells were cultured in 5% CO_2, atmospheric O_2 and at 37°C. ATDC5 mouse chondrocytes from 129 teratocarinoma AT805 are a model of cartilage differentiation was cultured in DMEM: Ham's F12 (1:1), GlutaMAX™ with L-glutamine (Gibco), 5% fetal Bovine Serum (FBS) (Gibco), 100 I.U/ penicillin and 100µg/mL streptomycin. To push ATDC5s into differentiation they were cultured in the presence of 1% insulin, transferrin, and selenium

(Insulin-Transferrin-Selenium-Sodium Pyruvate, ITS-A) (Gibco) for 28 days (Atsumi et al., 1990). The ATDC5 cells retain a fibroblast like morphology until differentiation in which they form cartilage like nodules and depose a matrix. The T/C-28-I2 are immortalised human costal cartilage from primary culture (Goldring et al., 1994). Cultured in DMEM: Ham's F12 (1:1) with L-glutamine, 10% FBS (Gibco), 100 I.U/ penicillin and 100µg/mL streptomycin. A fibroblast cell line was used to determine the activation of enhancers in a different cell type. These were the murine embryonic fibroblast, NIH3T3, maintained in DMEM high glucose (4500mg/L) plus L-Glutamine (Gibco), 10% FBS (Gibco), 100 I.U/ penicillin and 100µg/mL streptomycin. For transfections FBS was replaced with 2% TCM® (MP biomedicals) define serum replacement as it contains no growth factors, steroid hormones or extraneous proteins that could interfere with cytokine treatment.

3.9.1. qPCR to determine chondrocyte phenotype of ATDC5 cells

ATDC5 were seeded to a density of $2.5x10^4$ in 48 well cell bottom tissue culture plates in differentiation media for 48hours prior to RNA extraction

3.9.1.1. TRIzol® based RNA extraction

RNA was extracted using TRIzol® RNA isolation reagent (Invitrogen). Media was removed from each of the wells and washed once with room temperature PBS. All PBS was removed and 200µL of TRIzol® was added to each well and mixed by pipetting and incubated at room temperature for no more than 5 minutes. The solution was then placed into RNase and DNase free 1.5mL centrifuge tubes, 200µL of chloroform was added to the tube and vigorously shaken by hand for 15 seconds and incubated at room temperature for 3 minutes. Subsequently, centrifuged at 12,000g at 4°C for 15 minutes. The clear aqueous phase was removed by pipetting and placed into a fresh RNase and DNase free 1.5mL centrifuge tube, 10µg of GlycoBlue™ (Invitrogen) and 500µL of isopropanol was added to the solution and inverted twice and left at room temperature for 10 minutes. Centrifuged at 12,000g for 10 minutes at 4°C and the supernatant was discarded. 1mL of ice cold 75% ethanol was then added and vortexed for 2 seconds before centrifugation at 7500g for 5 minutes at 4°C to wash, the supernatant was discarded and the RNA pellet

was left to air dry for 10 minutes. 50µL of RNase-free water was added to the RNA pellet, mixed by pipetting and placed at 55°C for 15 minutes. RNA concentration and purity was determined by UV spectroscopy.

3.9.1.2. cDNA synthesis

RNA was diluted to 100ng/µL and used for cDNA synthesis using the High-Capacity cDNA reverse transcription kit (Applied Biosystems). Reactions were set up in 20µL reactions with 2µL 10x RT buffer, 0.8µL 25x dNTP mix, 2µL RT random primers, 1µL MultiScribe™ reverse transcriptase, 4.2µL RNase-free water and 10µL RNA (100ng/µL). The reaction was then carried out in a thermocycler with the parameters: 25°C for 10 minutes, 37°C for 120 minutes, 85°C for 5minutes and held at 4°C.

3.9.1.3. Real time PCR

qPCR was carried out using the SensiMix™ SYBR® Hi-ROX kit (Bioline) for *MMP13, Sox9, Runx2, Col1a1* and *Col2a1*. cDNA was diluted 1:5 and qPCR reactions were carried out in 20µL reaction consisting of 10µL 2x SensiMix™, 0.6µL of each of the forward and reverse primer (listed in table below), 6.8µL nuclease free water and 2µL of diluted cDNA for each sample. Each reaction was performed in triplicates per sample. A three-step with melt reaction was performed on the Corbett (now Qiagen) Rotor Gene RG-6000, initial step of 95°C for 10 minutes, then 40 cycles of 95°C for 15 seconds, 60°C for 10 seconds, 72°C for 30 seconds and acquire on green and orange (for ROX loading control) channel. A melt curve was generated at the end of the run to ensure a single product had been amplified. Housekeeping gene and target genes were performed in the same runs to avoid discrepancies. Gene expression analysis using TaqMan® assays was used for *Acan* (mm00545794_m1) with the house keepers hypoxanthine guanine phosphoribosyl transferase (*Hprt*) (mm00446968_m1) and *Rplp0* (Mm01974474_gH) in a mix containing 1X TaqMan® Universal PRC mix (ABI), 1XTaqMan® gene Expression Assay , 100ng of cDNA in a final volume of 20µL. The cycling conditions 50° for 2 minutes, 95°C for 10 minutes, 95° for 15 seconds and 60° for 1 minute for 40 cycles detection on Green and Orange. Samples were analysed in three technical replicates with NTCs to allow normalisation.

Results were annotated on the Rotor-Gene 6000 series software version 1.7. The housekeeping gene used for *Col1a1* was *18S*, for *Col2a1* and *MMP13* GAPDH, for *Sox9* and *Runx2* RPLP0. ΔCT was calculated for each gene after normalisation to the housekeeping gene using:

$$= 2^{-(average\ CT\ of\ gene\ of\ interest - average\ CT\ of\ housekeeping\ gene)}$$

Primer name	**Sequence 5' – 3'**
mSox9 F	GAC AAG CGG AGG CCG AA
mSox9 R	CCA GCT TGC ACG TCG GTT
mRunx2 F	CCC AGC CAC CTT TAC CTA CA
mRunx2 R	TAT GGA GTG CTG CTG GTC TG
mCol1a1 F	CAT GTT CAG CTT TGT GGA CCT
mCol1a1 R	GCA GCT GAC TTC AGG GAT GT
mCol2a1 F	TGG GTG TTC TAT TTA TTT ATT GTC TTC CT
mCol2a1 R	GCG TTG GAC TCA CAC CAG TTA GT
m18S F	GGA AAG CAG ACA TCG ACC TCA
m18S R	AGT TCT CCA GCC CTC TTG GT
mGAPDH F	GGG CTC ATG ACC ACA GTC CAT GC
mGAPDH R	CCT TGC CCA CAG CCT TGG CA
mRPLP0 F	ATG GGT ACA AGC GCG TCC TG
mRPLP0 R	GCC TTG ACC TTT TCA GTA AG

To determine the efficiency of primer pairs the following calculation was used and the primers were initially examined using standard curves:

$$Efficiency = \left(10^{-\frac{1}{slope}} - 1\right) \times 100$$

3.9.2. Lipofectamine mediated transfection

ATDC5 were seeded to a density of $2.5x10^4$ cells/well, T/C 28 I2 at $3.0x10^4$ cells/well and NIH3T3 at $1.75x10^4$ cells/ well in 48 well clear bottom tissue culture treated plates overnight (around 12 hours). Transfection were only conducted when cells were around 80% confluent and completed using Lipofectamine® 2000 (Invitrogen). Cells were co-transfected in duplicates with the enhancer constructs and *Renilla* luciferase control vector, pRL-TK (Promega) to allow normalisation of transfection efficacy. Blanks, pGl3 vector with no inserts and pRL-TK alone were included in each experiment as controls

For each experiment the following reaction was set up:

Tube 1	Tube 2
80 µL Opti-MEM® Medium	80 µL Opti-MEM® Medium
2µg DNA	3µL Lipofectamine® 2000
120ng pRL-TK vector	

After 5 minutes of incubation at room temperature 80µL from tube 2 was added to tube 1, mixed by flicking and incubated at room temperature for 25 minutes. 40µL of the mix was added to each well so the final concentration of DNA was 500ng with 0.75µL of lipofectamine and 30ng of TK- vector. Cells were assayed after 48 hours or treated with catabolic agents before assays.

3.9.3. Induction of catabolic events in cells

Oncostatin M (OSM) produces mild catabolic events in chondrocytes through JAK3 and suppresses *Acan* expression and elevates the effects of IL-1 in human, rat and mouse chondrocytes (Goldring and Otero, 2011) and IL-1β has been implicated in OA, by increasing the enzymes which break down the matrix. To examine if these two cytokines effects the expression of the *Acan* enhancer we treated transfected cells with these agents. Human oncostatin M (hOSM) (Cell Signalling) was diluted in PBS to a concentration of 1µg/mL and human IL-1β (Cell Signalling) with carrier was diluted in PBS to the same

concentration. 24hours after cells were transfected with luciferase based vectors the media was replaced with media containing either 10ng/mL of hOSM or 10ng/mL of hIL-1β for 24hours before luciferase assay.

3.9.4. Luciferase cell assay

Luciferase assays were conducted using the Dual-Luciferase® Reporter Assay system (Promega). Reagents were prepared fresh for each experiment, with the exception of the Luciferase Assay Reagent II (LARII). Passive lysis buffer was prepared by adding 1 volume of 5X concentrate into 4 volumes of distilled water and inverted to mix. (LARII) was prepared by adding 10mL of Luciferase Assay Buffer II to the lyophilised luciferase assay substrate and stored in aliquots of 1mL at -80°C 1 volume of Stop & Glo® substrate was added to 50 volumes of Stop & Glo® buffer and inverted to mix. Cells were allowed to equilibrate to room temperature for 10 minutes and then 65µL of Passive lysis buffer was added to each well and incubated at room temperature for 15 minutes while rocking. 50µL of LARII was added to each well of a white 96-well plate and a background reading was read, 10µL of cell lysis was transferred into the plate and mixed by pipetting, within 10 minutes the plate was read in the GloMax®-multi detection dual injector system equipped with a 96-plate holder and luminescence module read at 0.3 seconds per well. 50µL of Stop & Glo® Reagent was added to each well, mixed before the *Renilla* luminescence was read. The luminescence was normalised to the control vector and the relative response ratio (RRR) was calculated using the following formula:

$$RRR = \frac{(experimental\ sample\ ratio) - (negative\ control\ ratio)}{(positive\ control\ ratio) - (negative\ control\ ratio)}$$

3.9.5. Statistical analysis

Paired T-test was conducted to determine if there was a fold change between samples and treatments. A P value of <0.05 was considered significant.

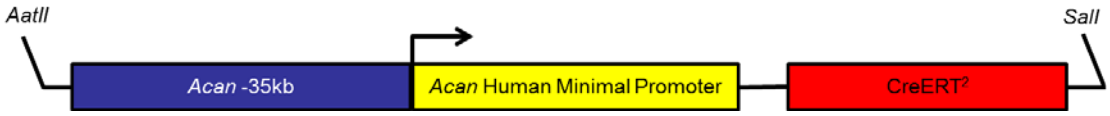
3.10. Analysis of Acan -35kb Cre recombinase mouse

3.10.1. Preparation of tamoxifen for IP injection

Tamoxifen (Sigma) was prepared by dissolving in 200 proof absolute ethanol to a concentration of 100mg/mL by vortexing to create stock solution. Stock was emulsified in corn oil (Sigma) to 10mg/mL. Prior to injection the emulsified solution was sonicated and kept at room temperature, fresh batches were made before each injection.

3.10.2. Generation of *Acan* -35kb Cre mouse

The *Acan* -35kb cre was created by Grace Chu, formally of the University of Liverpool/ Kennedy Institute of Rheumatology, Oxford, the construct was named pChu9 after the creator. The *Acan* -35kb enhancer was blunt end ligated to the human *Acan* promoter. The pCreERT^2 vector (Feil et al., 1997) was digested with *PvuII* and *BglII* and blunted using *T4* DNA polymerase (NEB) following manufacturers instruction and dephosphorylated using FastAP thermosensitive alkaline phosphatase (Thermo Scientific). Blunt end ligation was conducted at a 4:1 ratio overnight at 16°C to create the pChu9 vector. To prepare for oocyte microinjection the pChu9 was digested with *SalI* and *AatII* to linearised. Four lines of pChu9 were created: pChu91, pChu92, pChu93 and pChu94 for analysis.



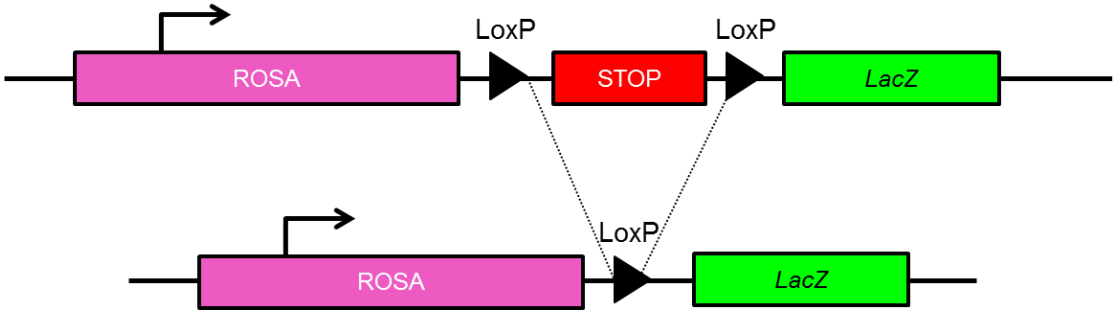
3.10.3. Genotyping of *Acan* -35kb cre

Genotyping was carried out as previously described using primers specific to cre recombinase and *LacZ* with internal.

Primer Name	Sequence 5' – 3'	Species	Annealing Tm (°C)	Target	Product Size
HELp010	GCA TTA CCG GTC GAT GCA ACG AGT GAT GAG	Mouse	68	*Cre*	
HELp011	GAG TGA ACG AAC CTG GTC GAA ATC AGT GCG	Mouse	68	*Cre*	389bp

3.10.4. Analysis of *Acan* Cre x R26R reporter mouse

To determine the best transgenic line to take forward we utilised the *Gtrosa26*^{tm1Sor} (R26R) reporter mice. The R26R allows to detection of a cre expression pattern, as the activation of the cre removes the "STOP" codon that are flanked by *LoxP* sites. The recombination allows the expression of the *LacZ* gene (Soriano, 1999). Male pChu94 mice were time mated with a female R26R mouse and 1mg of tamoxifen was administered by intraperitoneal injection (IP) on day 13 of pregnancy with half dose of progesterone (Sigma) (up to 2ng) to reduce abortion rates, injections were carried out by Dr Ke Liu. Females were sacrificed and embryos collected at E15.5 and stained for *LacZ*. Adult pChu9 x R26R mice were injected with 3 doses of 1mg tamoxifen at 8 weeks from each line were sacrificed 8 weeks post injection to determine adult expression of each line by *LacZ* staining.



4. RESULTS

4.1. Characterisation of chondrocyte specific Acan enhancers

4.1.1. In silico identification of possible Acan enhancers using ENCODE

To establish regions to examine for *cis*-acting elements of *Acan* we first screened the TAD of the gene with a search range of -150kb to +65kb of the TSS (chr7:86,069,981-86,262,128 in mouse), as *Acan* appears to contain a dessert of non-coding region upstream of the promoter it was the best region to start. Using the mm9 mouse accession, as it was the most complete and readily provided the data tracks for histone modifications and DNaseI hypersensitivity in limb only. This in contrast with the earlier and the latest accession of the human and mouse genomes on the genome, where user input were more heavily relied on, these accession would be used in later analysis to strengthen our predictions. Identification by eye was then carried out or a parametric linkage analysis, LOD was used. After comparison we deemed by eye was the quickest and generated the same data.

LOD scoring ranked regions where the vertebrate Phastcons are high, with intersections of overlap with H3K4me1, H3K27ac and absence of H3K3me3 in the limb and limited to no overlap and a score greater than or equal to 300. With LOD analysis 5 regions upstream of *Acan* showed promise for enhancer activity, the -115kb, -100kb, -87kb, -35kb and the already known enhancer at -10kb (Figure 17). Based on the histone modification peaks we predicted 16 possible enhancers that lie intronic and intergenic of *Acan* (Figure 18). We took a selection of possible regions to examine further, the -35kb and -87kb contained a strong level of H3K27ac and H3K4me1 localisation and low H3K4me3 indicating the presence of an actively open chromatin at E14.5, high LOD score but had low limb specific DNaseI hypersensitivity. The next to be considered was the -65kb region as it contained defined peaks of H3K4me1, H3K427ac and both hind and fore limb E13.5 DNaseI hypersensitivity indicative of a active enhancer. Although there is a strong mark of

H3K4me3 commonly associated with promoters it has been noted that H3K4me3 is deposited at enhancers where PolII is highly associated and for this reason the LOD analysis excluded this region (Chen et al., 2015, Pekowska et al., 2011). Additionally, we looked at 3 intronic sequences that did not get flagged by the LOD analysis, the +26kb because of the histone marks and limb specific DNaseI, and in the final two introns contained high conservation which indicated a highly important region in the gene and worth pursuing although that had little or no histone modifications in the limb. We utilised the function of liftover on the genome browser to examine the regions in humans to see if there was an agreement between the two species and to allow further work to be done in understanding the enhancer's role in humans. The hg19 accession provides information on the chromatin that are not available in mice, however lacks limb specific data tracks which may mean these predicted enhancers may not have chondrocyte specificity. On the other hand these data sets offers histone modifications, chromatin interactions analysis by paired- end tags (ChIA-PET), chromatin state using ChIP-seq a Hidden Markov Model (HMM) in human cell types.

Table 5: Cells used in the ENCODE project

Cell name	**Cell type**
GM12878	Lymphoblast
H1-hESC	Human embryonic stem cells
K562	Mesoderm linage: bone marrow
HepG2	Hepatocytes
HUVEC	Endothelial
HMEC	Dermal microvascular endothelium
HSMM	Myoblasts
NHEK	Epidermal keratinocytes
NHLF	Lung fibroblast
HeLaS3	Epithelial

This allows us to assess the area and our predicted enhancers for their function in other cells types and to determine if these enhancers have predicted functions in humans (Figure 19). The chromatin state by HMM indicated that in cells other than chondrocytes that our enhancers are on heterochromatins or inaccessible by the experiments. The +26kb appears to act as an insulator and interacts with the repressor CTCF in ChIA-PET. The -35kb is repressed in most cell types but in HMEC and HSMM where it is a weak enhancer and a poised promoter in HUVECS and h1-hESC (Figure 19). This requires further validation but builds on a picture of how enhancers drive their cell type specificity which will be discussed in detail later.

We examined the +60, +55, +26, -35, -65 and -87kb by cloning 100bp either side of the histone peaks into the HSP68/*LacZ*/GW and creating transgenic mice, the names of each enhancers are based on an estimation from UCSC genome browser and does not represent the exact distance from the TSS. Enhancer sizes ranged from 2.5kb to 700bp initially before regions were shorten to a minimum of 400bps. The exact locations and size for mouse and humans are given in Table 6.

Table 6: Location of possible aggrecan enhancers in the mouse and human accessions

Enhancer	**Mouse (mm9)**			**Human (hg19)**		
	Coordinates	Size (bp)	Actual distance (kb)	Coordinates	Size (bp)	Actual distance (kb)
+60kb	**chr7:86,258,701-86,259,111**	**411**	**+60.3**	**chr15:89,417,233-894,17,661**	**429**	**+70.6**
+55kb	**chr7:86,251,764-86,254,274**	**2511**	**+53.3**	**chr15:89,410,372-89,412,708**	**2337**	**+64.0**
+26kb	**chr7:86,224,417-86,226,396**	**2015**	**+28.0**	**chr15:89,376,262-89,378,225**	**1964**	**+31.6**

+26kb Small	*chr7:86,224,823-86,225,626*	*804*	*+26.0*	*chr15:89,376,621-89,377,479*	*859*	*+29.9*
+3kb	chr7:86,201,124-86,202,499	1376	+2.8	chr15:89,350,322-89,351,819	1498	+3.6
***-10kb**	**chr7:86,188,658-86,189,016**	**504**	**-9.7**	**chr15:89,335,961-89,336,323**	**363**	**-10.7**
-27kb	chr7:86,171,136-86,172,814	1679	-27.2	chr15:89,314,768-89,316,511	1744	-31.9
-35kb	**chr7:86,165,553-86,167,458**	**1906**	**-30.9**	**chr15:89,311,396-89,313,708**	**2313**	**-33.0**
-35kb Small	**chr7:86,166,240-86,166,651**	**412**	**-31.7**	**chr15:89,312,808-89,313,168**	**361**	**-33.5**
-60kb	chr7:86,138,571-86,139,868	1298	-59.8	chr15:89,291,111-89,292,883	1773	-55.6
-65kb	**chr7:86,134,232-86,136,363**	**2132**	**-62.0**	**chr15:89,283,882-89,286,567**	**2686**	**-60.1**
-68kb	*chr7:86,130,851-86,132,182*	*1332*	*-67.5*	*chr15:89,282,199-89,283,340*	*1142*	
-87kb	**chr7:86,115,696-86,118,311**	**2616**	**-80.1**	**chr15:89,259,357-89,262,271**	**2915**	**-84.4**

-87kb Small	**chr7:86,116,387-86,117,140**	**754**	**-81.2**	**chr15:89,260,172-89,260,980**	**809**	**-85.7**
-90kb	chr7:86,108,776-86,111,305	2530	-89.6	chr15:89,251,996-89,256,471	4476	-94.7
-100kb	chr7:86,097,840-86,100,140	2301	-100.5	chr15:89,238,296-89,240,649	2354	-108.4
-108kb	*chr7:86,090,087-86,091,743*	*1657*	*-108.3*	*chr15:89,233,127-89,234,795*	*1669*	*-113.5*
-115kb	*chr7:86,083,718-86,084,710*	*993*	*-114.7*	*chr15:89,224,742-89,225,738*	*997*	*-121.9*
-128kb	*chr7:86,070,048-86,071,779*	*1732*	*-128.3*	*chr15:89,202,247-89,205,439*	*3193*	*-144.4*

Genomic co-ordinates for enhancer regions and their size in the mouse and human accessions mm9 and hg19 respectively. The actual distance from Acan transcriptional start site (TSS) is calculated by the 5' co-ordinate of the enhancer to the 5' of the TSS in kilobases. ***Bold*** *are the enhancers tested in vivo, italics are the enhancers that reporter constructs were created but never examined.*

We chose to use Gateway® based cloning due to the high throughput and efficient generation of construct as well as the ease and availability of a destination vector with a silent promoter to allow assessment of enhancer activity (the HSP68 minimal promoter). One of the major pitfalls of using this system is that we did not establish a baseline of the HSP68 construct with no enhancers attached and we failed to examine the mouse or

human minimal promoter working alone or alongside the enhancers to establish the endogenous pattern of expression. But not withstanding we were able to compare each enhancer to each other. The *LacZ* used in the constructs is a bacterial reporter from *Escherichia coli* that encodes the enzyme β-galactosidase (β-gal). β-gal is able to catalyse the hydrolysis of β-galactosides into monosaccharides., When introduced to the β-galactoside 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside (X-gal) it has its benefit of acting as a reporter for gene activity. B-gal is able to hydrolyse the X-gal substrate into galactose and 5-bromo-4-chloro-3-hydroxyindole, then two of the second product is then oxidised into the insoluble 5.5'-dibromo-4,4'-dichloro-indigo precipitate.Which, as the name suggests, is a indigo substrate that is easy to visualise (Anson and Limberis, 2004). Both dimerization and oxidation steps can only occur in the presence of ferric and ferrous ions that acts as electron acceptors (Lojda, 1970).In transgenic mice the blue staining indicates the activity of the transgene and therefore the expression pattern of the enhancer.

Initially, for screening we used founder embryos in "transient experiments" to assess the function of the *cis*- acting elements due to the higher throughput compared to generating lines. However, we would have to re-inject the constructs if we wanted to analyse the regions in more detail. We collected founders at E15.5 which falls within the age of the histone modifications peaks found on the ENCODE tracks (E14.5) due to developmental rates being variable. Collections of founders at E16.5 or greater was avoided due to the keratinisation of the skin that reduced the penetration of the substrate. We also could not account for the insertion sites and copy number introduced into the genome therefore the possibility of site of integration (SOI) was always present. This was eliminated or reduced by identification of common expression patterns within the same litter or creation of lines for some of the enhancers.

β-gal activity is pH dependant and endogenous activity within certain tissues such as heart, skeletal muscle, brain, kidney and spleen can occur. Exogenous expression can be differentiated by using a stain solution of around pH 8.0 (Tenu et al., 1971, Weiss et al.,

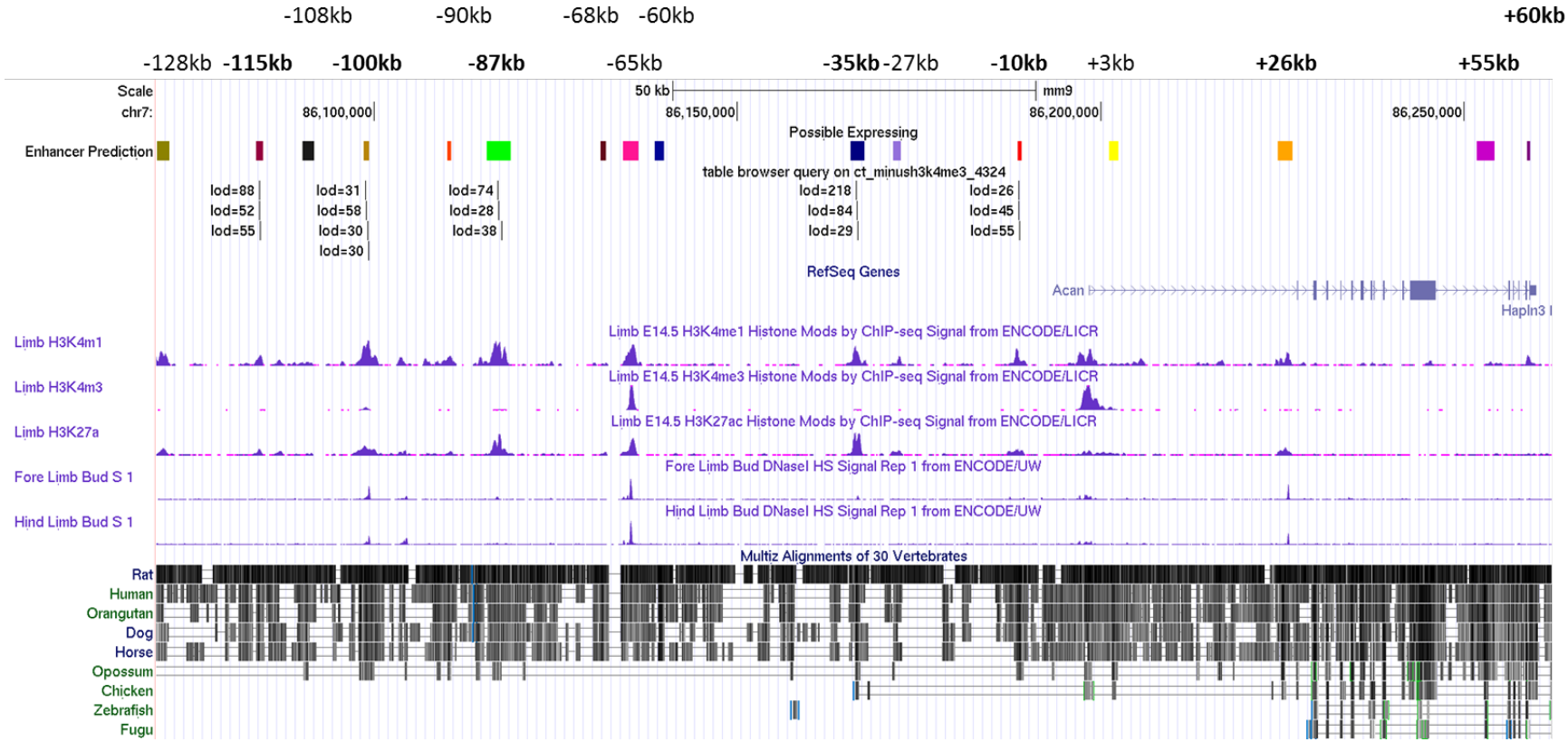
1999). Furthermore, endogenous *E.coli* and activity from metallic contamination is possible but can be differentiated by observing bilateral and symmetrical expression.

Interestingly, all the enhancer regions examined expressed in transgenic mice at E15.5 and these were added as custom tracks for mouse and human on the UCSC browser for comparison (Figure 18 and Figure 19). The results of each enhancers will be discussed further and are grouped into:

- +26kb and -35kb enhancers
- -65kb and -87kb enhancers
- -55kb and +60kb enhancers

To discuss the properties and their roles in regulating *Acan*.

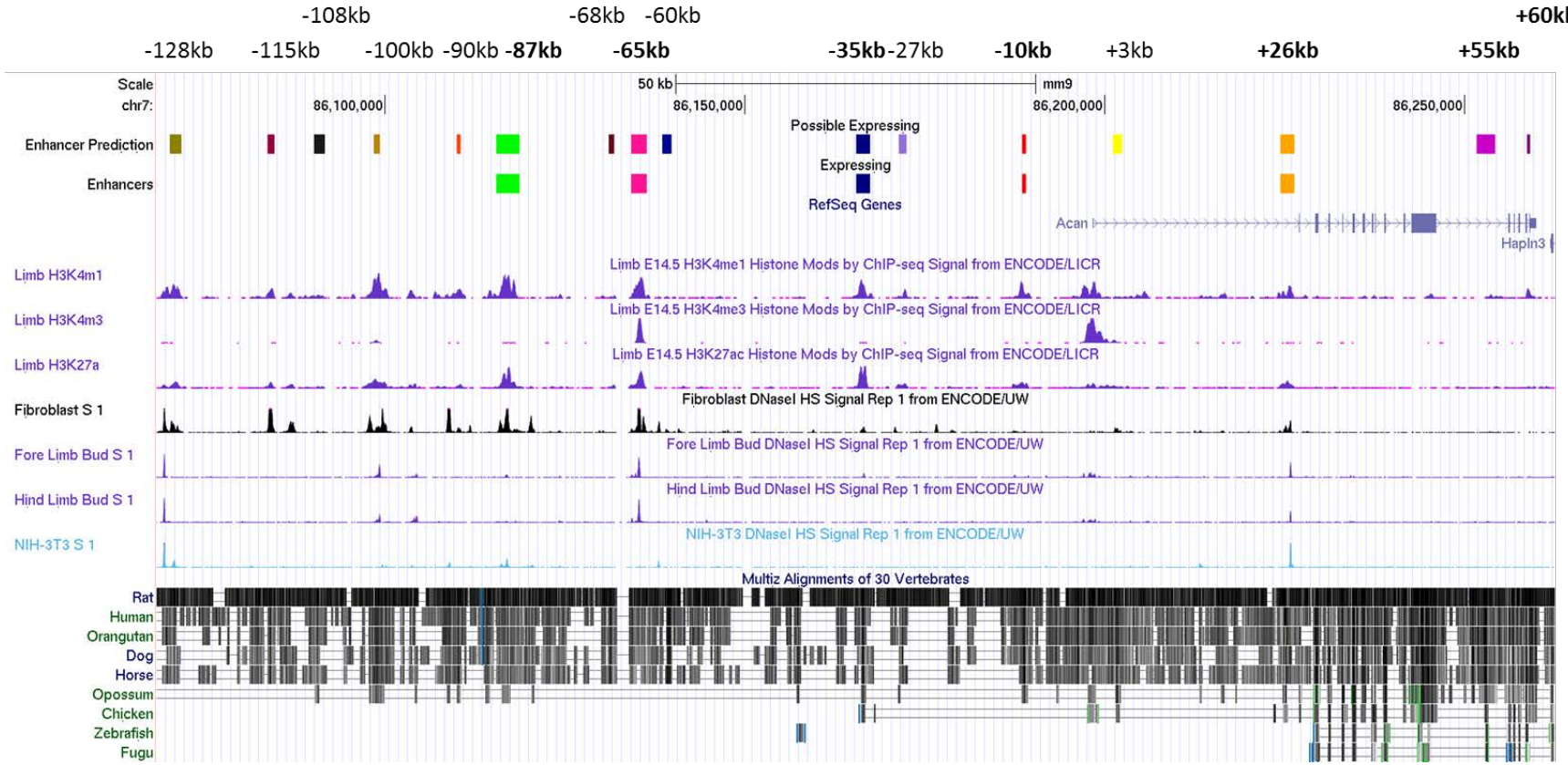
Figure 17: LOD analysis of *Acan* locus for possible enhancer compared to user determined regions



Read out on UCSC genome browser compared to regions identified by user and regions identified using LOD scoring on the mouse mm9 accession. The

tracks from top to bottom. "Enhancer prediction" are the user predicted enhancers colour co-ordinated so region can be compared between accessions, in black with "lod=" are the LOD scores, the higher the number the stronger the prediction. The Limb histone and DNaseI hypersensitivity are in purple.

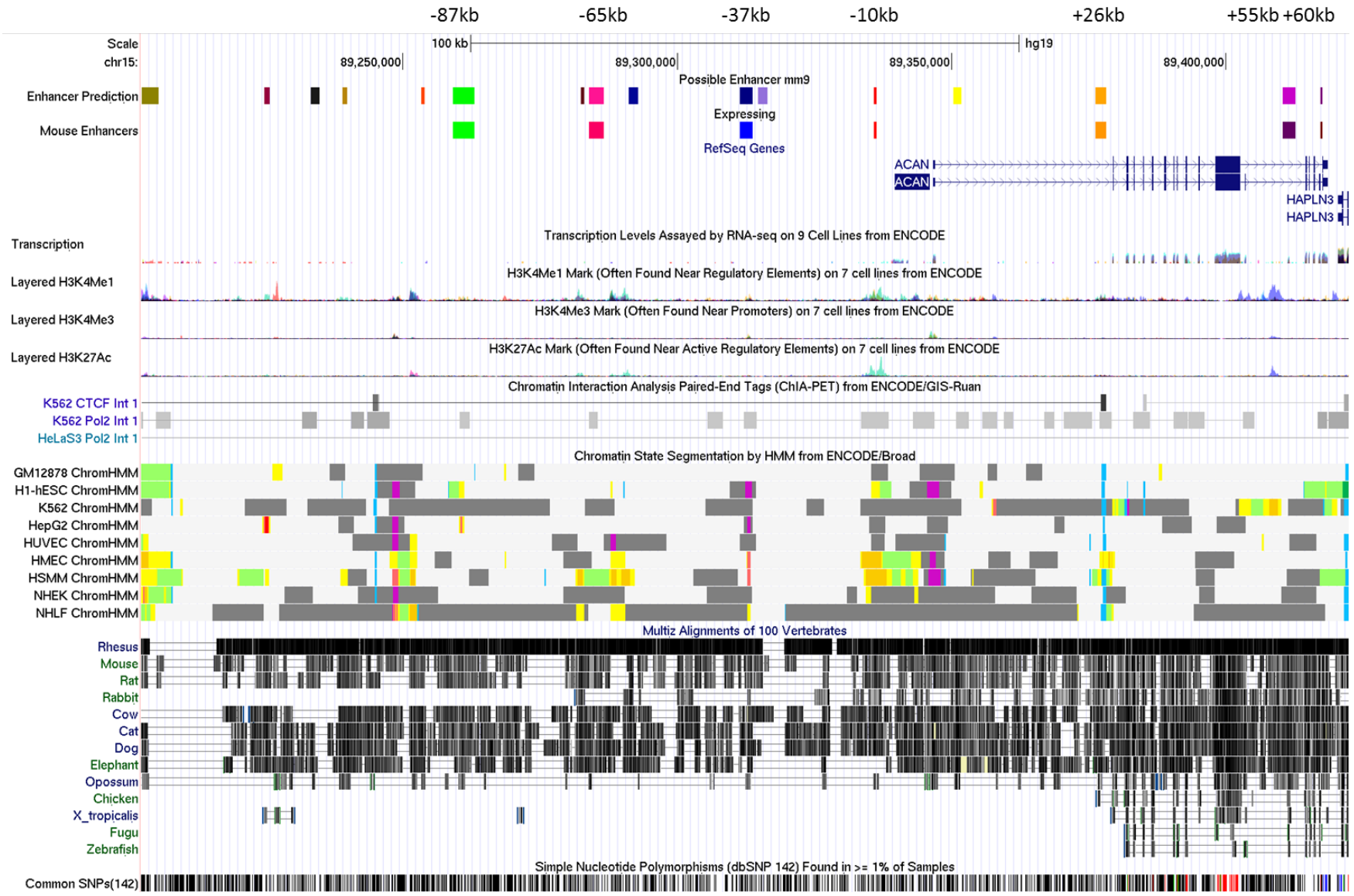
Figure 18: USCS limb specific histone modifications and identification of possible enhancers in *Acan*



Read out on UCSC genome browser compared to regions identified by user and regions tested using transgenic mice on the mouse mm9 accession. The tracks from top to bottom. "Enhancer prediction" are the user predicted enhancers colour co-ordinated so region can be

compared between accessions. "Enhancers" are the enhancers that express. The Limb histone and DNaseI hypersensitivity are in purple. The fibroblast DNaseI hypersensitivity is in black.

Figure 19: Liftover into human accession for comparison of predicted enhancers


-87kb
-65kb
-37kb
-10kb
+26kb
+55kb +60kb
Scale
100 kb
hg19
chr15:
89,250,000
89,300,000
89,350,000
89,400,000
Possible Enhancer mm9
Enhancer Prediction
Expressing
Mouse Enhancers
RefSeq Genes
ACAN
ACAN
HAPLN3
HAPLN3
Transcription
Transcription Levels Assayed by RNA-seq on 9 Cell Lines from ENCODE
Layered H3K4Me1
H3K4Me1 Mark (Often Found Near Regulatory Elements) on 7 cell lines from ENCODE
Layered H3K4Me3
H3K4Me3 Mark (Often Found Near Promoters) on 7 cell lines from ENCODE
Layered H3K27Ac
H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE
Chromatin Interaction Analysis Paired-End Tags (ChIA-PET) from ENCODE/GIS-Ruan
K562 CTCF Int 1
K562 Pol2 Int 1
HeLaS3 Pol2 Int 1
Chromatin State Segmentation by HMM from ENCODE/Broad
GM12878 ChromHMM
H1-hESC ChromHMM
K562 ChromHMM
HepG2 ChromHMM
HUVEC ChromHMM
HMEC ChromHMM
HSMM ChromHMM
NHEK ChromHMM
NHLF ChromHMM
Multiz Alignments of 100 Vertebrates
Rhesus
Mouse
Rat
Rabbit
Cow
Cat
Dog
Elephant
Opossum
Chicken
X_tropicalis
Fugu
Zebrafish
Simple Nucleotide Polymorphisms (dbSNP 142) Found in >= 1% of Samples
Common SNPs(142)

UCSC genome browser readout for the human hg19 accession. Mouse predicted enhancers were liftovered onto the accession and the enhancers which expressed are coloured. The histone modification tracks are layered from different cell types. Chromatin interactions with PolII and CTCF in the cells K562 and HeLa cells by ChIA-PET experiments are the tracks that follow. The coloured tracks "Chromatin state segmentations by HMM from ENCODE/Board" map the chromatin state for each region

4.1.2. The +26kb enhancer expresses in all chondrocytes at E15.5

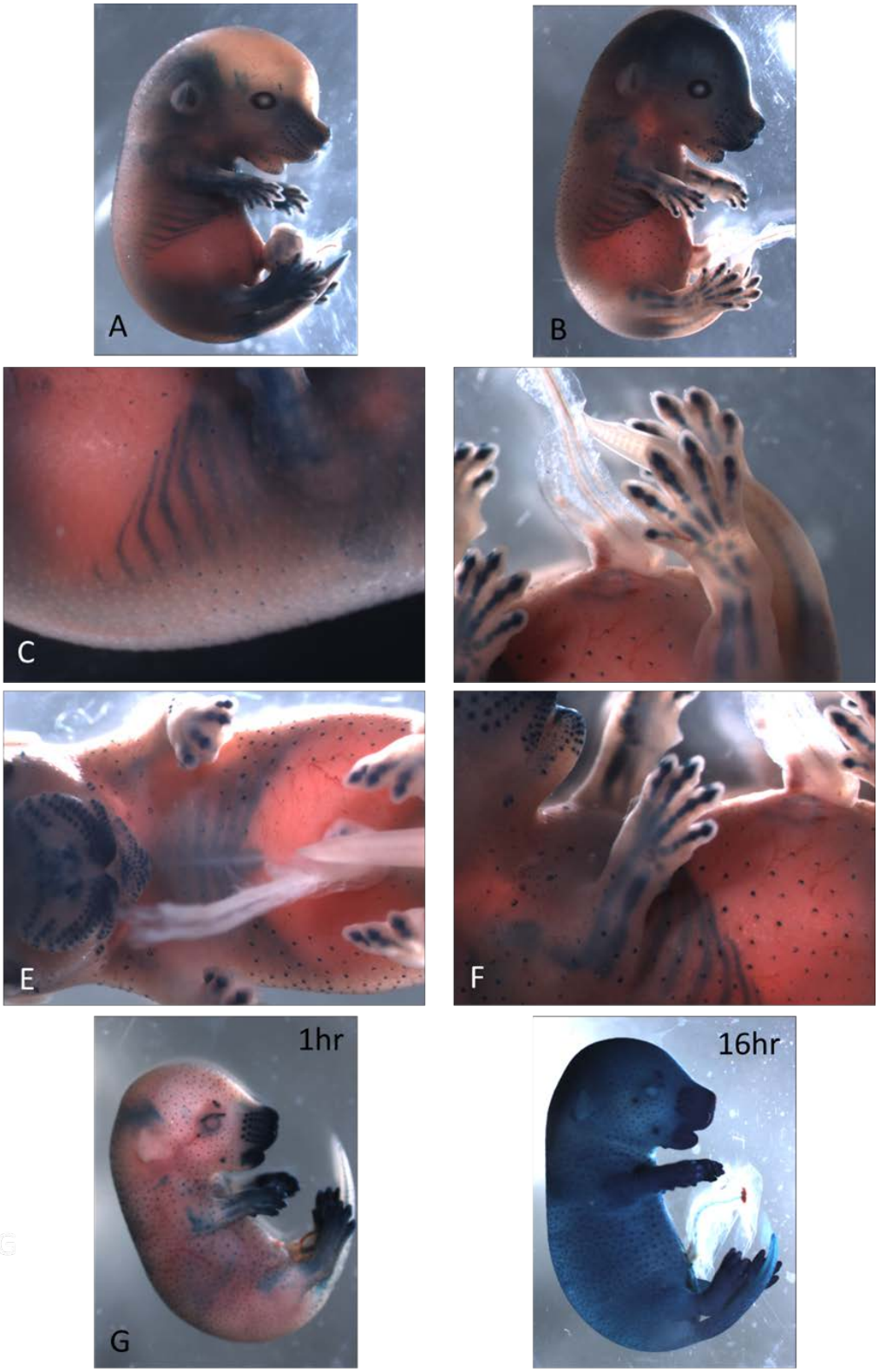
Intronic enhancers are not uncommon, they aid in the higher order of the chromatin and gene expression. The first intronic region of *Acan,* +26kb, was able to drive expression at E15.5 with consistent manifestations within three littermates, of which there were no other inserts, Table 7 gives an overview of the transgenic experiment and the percentage of expressing. This led us to believe that the expression was reproducible with little SOI or copy number effects. Whole mount images reveal strong activity in the growing limbs, digits, ribs and the tail following the characteristic expression of *Acan* in skeletal development but is not present in the craniofacial cartilage which appears absent when visually inspected. The enhancer is also able to drive expression in the vibrissae and hair follicles (Figure 20).

Table 7: Number of transgenic and expressing in transient experiments of the +26kb

Construct	**Expresser/ Transgenic (%)**	**Total embryos analysed in transient transfection**
+26kb	3/3 (100)	16

"Expresser" are the embryos which presented with β-gal expression. "Transgenic" were the embryos genotyped positive for LacZ and the percentage is the number of genotyped positive embryos compared to the expressers. The final column is the total embryos analysed in a litter

Figure 20: Whole mount images of the +26kb *Acan* enhancer at E15.5

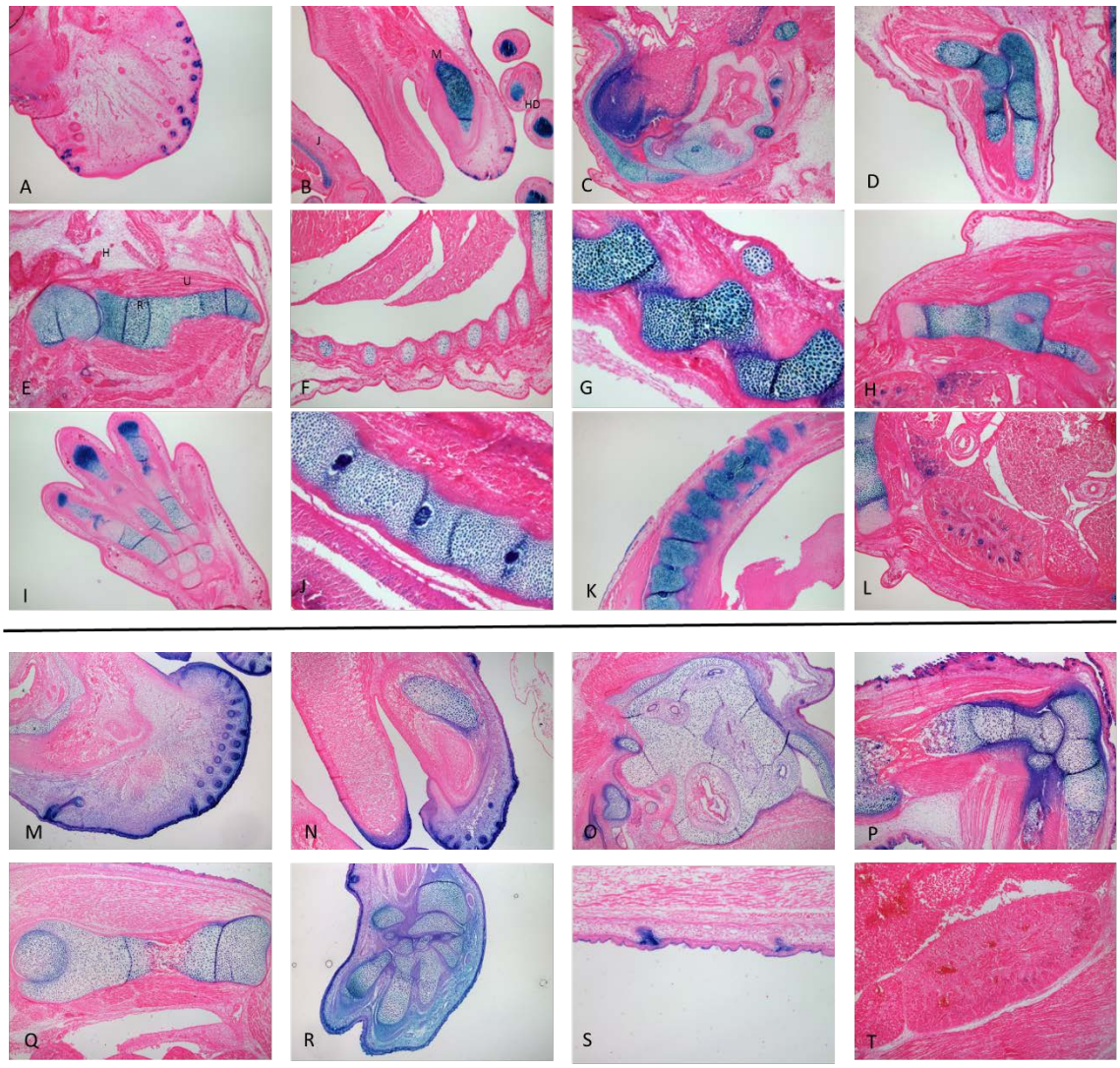


Whole mount images of the +26kb at E15.5. (A-B) full image of founder embryos. (F) higher magnification shows expression in the developing limb, humorous, ulna and radius and in the ribs (C and E). The bones in the limbs have high β-gal activity (D) with expression in the hair follicles. (E) coronal view of E15.5 embryo showing the vibrissae, ribs and sternum stained blue. G and H show another founder expression after 1hour and then 16hours showing intense skin expression but early skeletal staining.

Histologically with 6μM sections counter stained with eosin, which gives a good contrast between *LacZ* positive cells and unstained tissue, at E15.5 the +26kb element was able to drive tissue specific expression of the transgene in chondrocytes (Figure 21). The skeletal elements express the transgene seen in all chondrocytes including the hypertrophic cells in the forelimb, the occipital bones and ear, paw and ribs that follow normal expression of Acan expression (Figure 21) (Shibata et al., 2003). The expression of the +26kb is detected in the IVD, which arises from two different cell progenitors but do morphologically resemble chondrocytes (Kletsas et al., 2015, McCann et al., 2012). Staining is seen in both the outer fibrillary annulus fibrosis (AF) that surround the nucleus pulposus (NP).The detection of the transgene in the glomeruli of the kidney shares similar expression pattern to the Acan family member Versican (Rudnicki et al., 2012, Naso et al., 1995). This initially was usual as the glomeruli arise from the metanephric mesenchyme from the mesoderm which are different progenitors from the Acan expressing chondrocytes. Although Acan antibody staining has been localised to the apical epithelial tubules and basement membrane in the developing kidney at E15.5 (Lee et al., 2005). However, whenwe generated more founder lines, they showed similar chondrocyte staining but more intense skin staining and lacks kidney expression (Figure 21 M-T). These differences in expression between different founders could be a result of the random insertion of the transgene. Alternatively the expression may be in the capabilities of the enhancer but the sequence we used is to long (therefore, containing

silencing elements) or too short (missing elements), these suggestions are speculative and requires further confirmation.

Figure 21: Histological examination of the +26kb *Acan* enhancer at E15.5 shows chondrocyte expression



Representative histological sections of E15.5 embryos of the +26kb enhancer after overnight incubation with the X-gal substrate and counter stained with eosin. (A) Vibrissae of the superficial area present with blue staining. Chondrocytes are the primary expresser of the +26kb enhancer seen in the chondrocytes of the Meckel's cartilage [M],

jaw [J] and digits [HD] (B), skull and occipital bones (C), humerus [H], radius [R] and ulna [U] (D), humerus (E), ribs (F), vertebra (G), femur (H), metatarsal and phalanges of the hindpaw (I), the intervertebral disks of the lumber spine (J) and the tail (K). Interestingly, there appears to be specific staining in the glomeruli of the kidneys (L). Sections of another +26kb founder embryo show strong, non-consistent developing dermal skin staining as well as hair follicles (M-S) but chondrocyte staining in the ear (O), limb (P), femur (Q), paw (R) but no kidney expression is seen (T).

4.1.3. The -35kb enhancer expresses in chondrocytes during development

After the known enhancer (-10kb) (Han and Lefebvre, 2008) the first region in the non-coding DNA (intergenic) which we identify of significance is the -35kb. The development of the transient embryos generated high similarities between littermates. Therefore, to analyse further we generated lines from founder mice to determine any differences that could be as a result of copy number or SOI. We examined four lines initially lines 2, 8, 9 and 10 (Figure 22) and then took lines 2,8 and 9 forward for interrogation at different stages of development. The aim was to see if the enhancer was always on or switched on or off at different stages of cartilage and bone development up until post-natal maintenance.

At E15.5 whole mount show that the staining for the -35kb region was intense with a global localisation within skeletal elements, expressing in the axial and appendicular skeleton (Figure 22). The staining is complete with the entire cartilage structure being marked as opposed to the -10kb, which was confined to the core regions of the appendicular skeleton (Han and Lefebvre, 2008). When sectioned the embryos reveal chondrocyte specific deposition of the X-gal substrate (Figure 23). Expressing in the developing limb, ribs, Meckel's, the costal IVD and in the lumber IVD in both proliferating and hypertrophic chondrocytes. The costal and lumbar IVD differ by the lack of expression in the NP in the costal seen both in sagittal and coronal (Figure 23F and N). There are no expression in other tissues that are known to express *Acan* such as

the lungs and heart. This suggests that the -35kb enhancer is only capable of driving expression in cartilage and hints at the existence of non-chondrogenic enhancers.

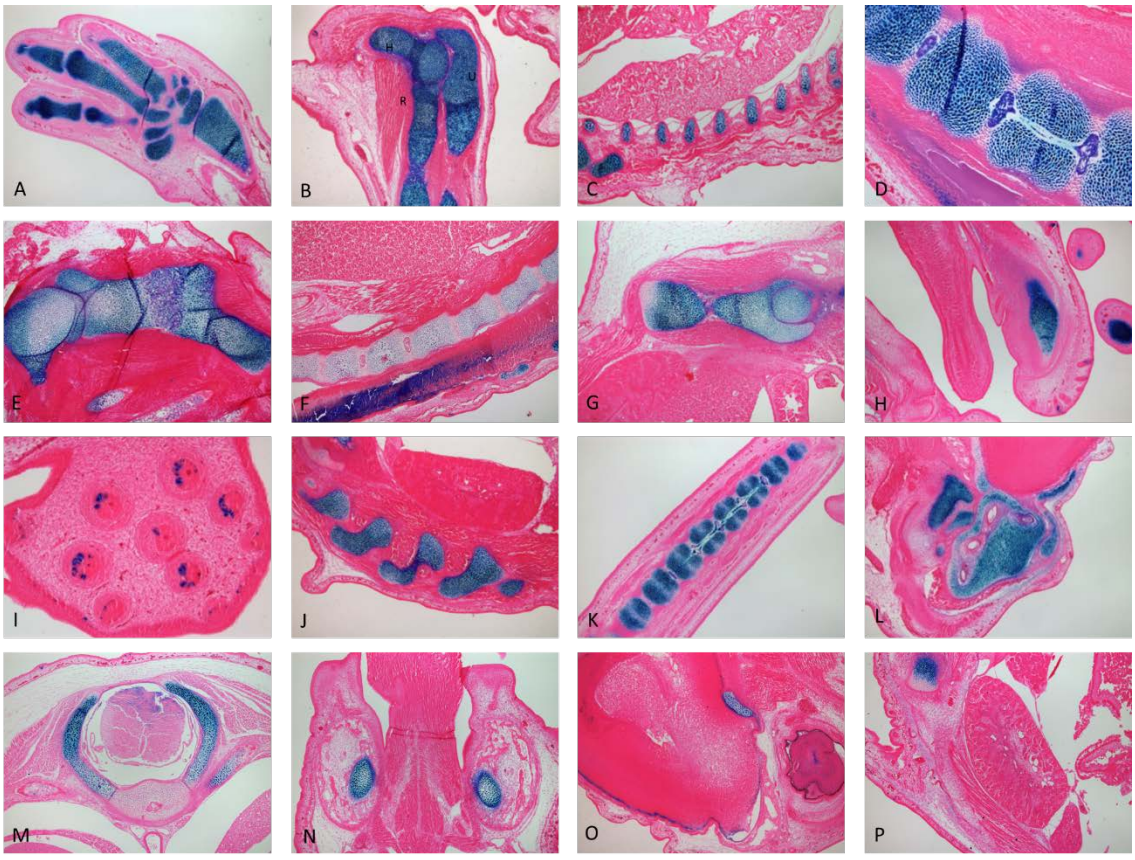
Table 10 provides an overview of the expressing regions of the transgene at E15.5.

To examine the temporal activity of the -35kb lines were examined at E13.5 and 8 weeks. At E12.5 the limbs are forming and expanding by E12.5, at E13.5 and between E13.5 and E15.5 there is a significant activation of genes as chondrocytes undergo hypertrophy and commitment to articular cartilage, osteoblast or other fates (Taher et al., 2011). Therefore, we wanted to see if there were any temporal dampening of the -35kb between these stages and into postnatal development examining at 8 weeks and 6 months of age (Pitsillides and Ashhurst, 2008, Decker et al., 2014a, Decker et al., 2014b).

Figure 22: Whole mount of the different -35kb embryos at E15.5

Whole mount images of the different -35kb lines at E15.5. (A) Line 2, (B) Line 8, (C) Line 9, (D) Line 10. (E) Coronal view of a representative of the spine showing activity of enhancer in the vertebra. (F) Sagittal view of the ribs and humerus. (G) ribs and vertebra sagittal. (H) coronal front view, showing tail, ribs and craniofacial staining. (I) complete skeletal elements staining in the forelimb and hindlimb (J). Clearing the embryos showed that in the -35kb there is strong skeletal staining in the spine and ribs seen in the whole mount (K-N) and staining covers the entirety of the long bones of the limbs (O-Q).

Figure 23: Histological examination of the -35kb *Acan* enhancer at E15.5



Histological analysis of the -35kb enhancer at E15.5. LacZ staining is present in the chondrocytes of the front paw's phalanges and metacarpals (A); the ulna [U], radius [R] and humerus [H] (B and E); in the ribs (C); the femur (G); the growing annulus fibrosus and nucleus pulposus of the lumbar spine and tail (D and K) but not present in the

nucleus pulposus of the costal spine (F); staining is present in the chondrocytes of Meckel's cartilage (H), the ear (L) and skull (O). Transverse sections show that the costal spine is only stained in cartilage tissue (M) and is present in the jaw (N). Staining is also noted in the vibrissae but absent from non- chondrocyte based tissues such as the eye (O), kidney (P), lungs (C) or tongue (H).

At E12.5 the transgene was clearly active in the developing limb cartilage and metatarsal-phalanges strongly between lines (Figure 25A-C, G-I). There was variability in craniofacial, somatic and brain expression of the transgene (Figure 25). Histologically, X-gal deposition is seen in the epaxial and hypaxial somites, the limbs and the developing jaw but not in the heart or lungs (Figure 25J-O). At 13.5 the limbs have progressed in development and closely resemble the expression observed at E15.5. Whole mount show manifestation of the transgene in the limbs, ribs and craniofacial (Figure 24A-F). Similar to the later stages of embryonic development we see chondrocyte specific expression but with no sign of decrease in any of the chondrocytes irrespective of their stage in maturation (Figure 24G-L).

Postnatally, at 8 weeks and 6 months the expression does not falter, here we present only the 8 weeks as there were no observational differences between 8 weeks and 6 months in terms of localisation and intensity of expression. The articular cartilage of the condyle and hip, the growth plates in the sternum and tail were clearly seen in whole mount after overnight incubation with X-gal (Figure 26). There was minor variability in the intensity in other cartilage types such as the xiphoid process and in the patella but no expression in non-cartilage tissues such as the ribs and kidney (Figure 26). The ureter in the kidney may have ectopic or endogenous expression at 8 weeks (Figure 26). The chondrocytes of the superficial region of articular cartilage (Figure 27C, F, G and K), the growth plates (Figure 27A, B and J) are the main expressers at 8 weeks of age as well as the superficial region of the costal cartilage. Other cells types in soft tissues such as in the lungs (Figure 27D) fail to express with the exception of the lumen of the aortic wall

(Figure 27H). The xiphoid showed little expression (Figure 27L) but the IVD showed strong expression in the end plates and in the AF but not in the NP (Figure 27B and I). This keeps the to the chondrocyte specificity of the -35kb seen at the earlier stages of development and marks out areas that overlap previously documented *Acan* expression although it does not mark all *Acan* expressing tissues (Fukada et al., 1999, Chambers et al., 2002, Melrose et al., 2008). The -35kb's ability to express in early development and into late adulthood suggests that it plays a key role in the regulation of its target gene. What needs to be established is its function, and if the knock-out of this region will have a deleterious effects.

Figure 24: Whole mount and histological images of the -35kb *Acan* enhancer at E12.5

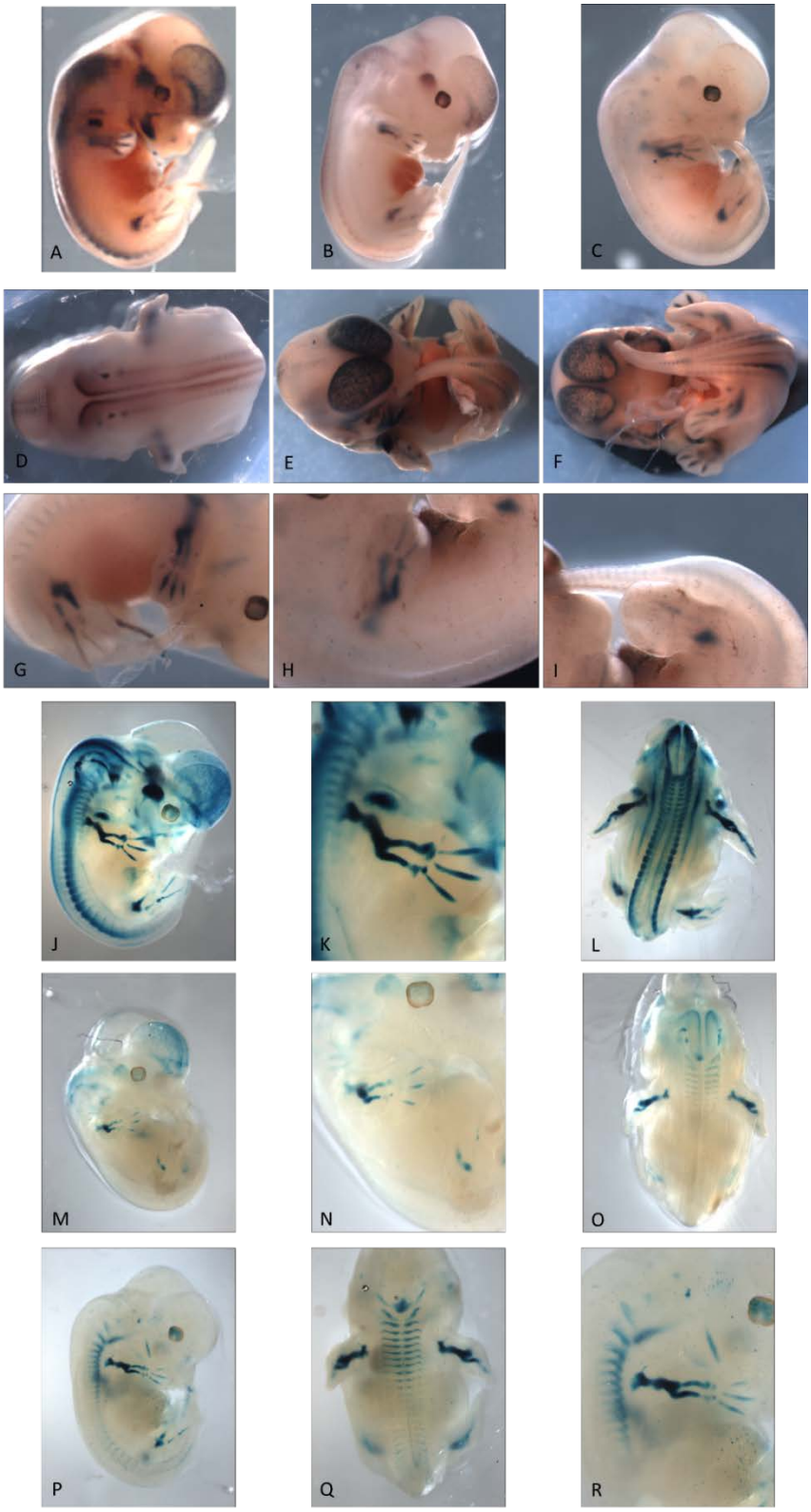
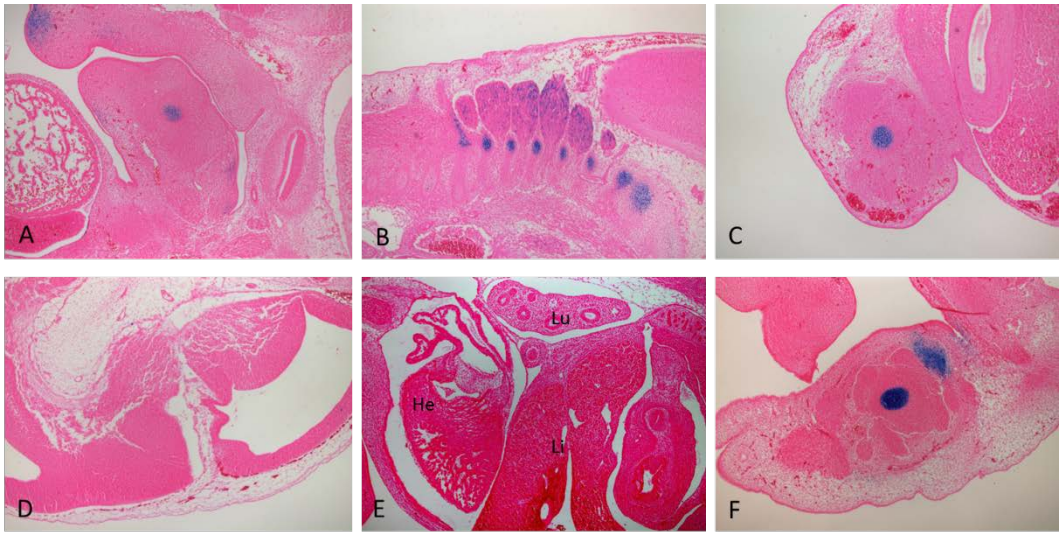
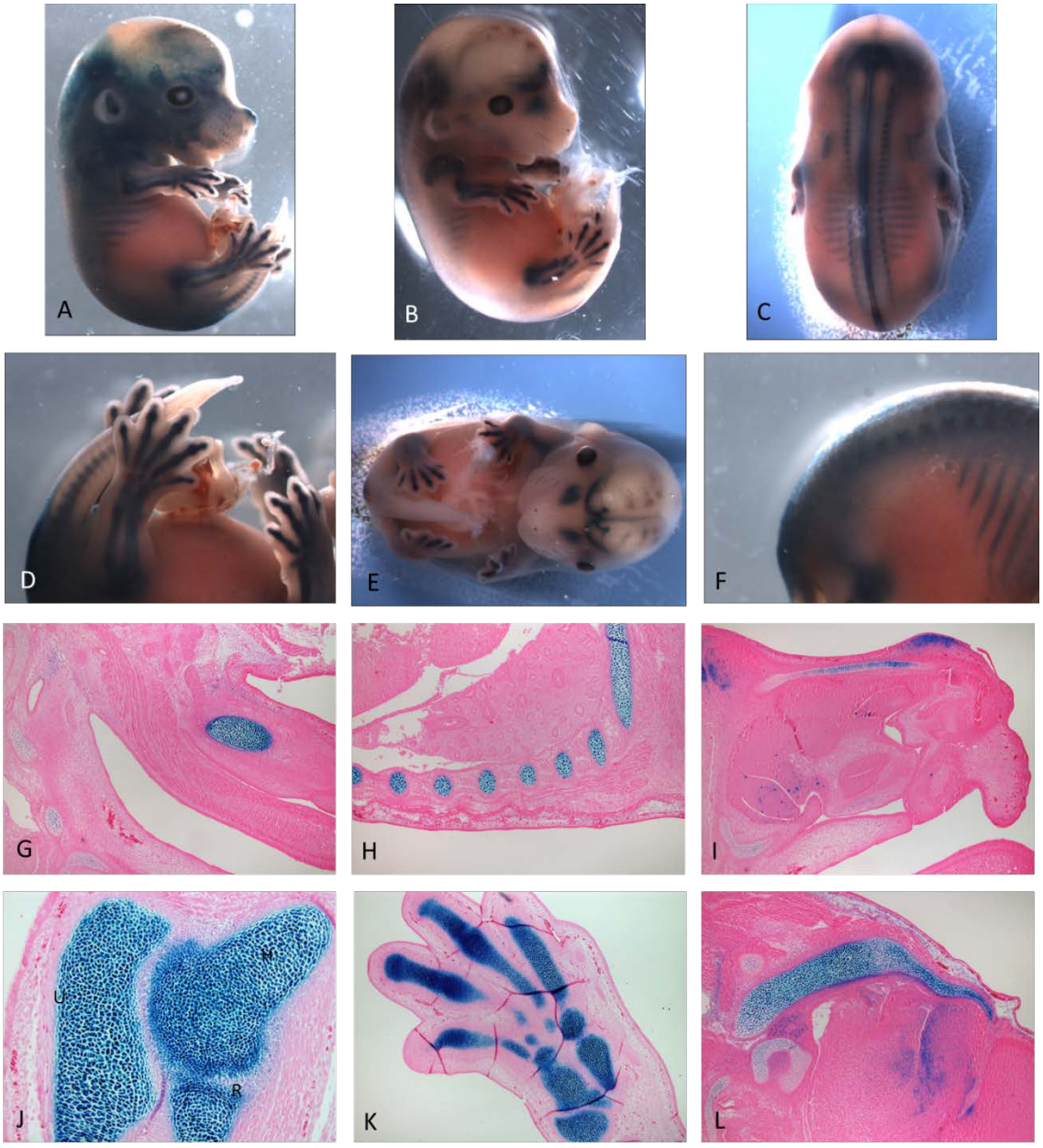


Figure 25: Histological examination of the -35kb enhancer at E12.5



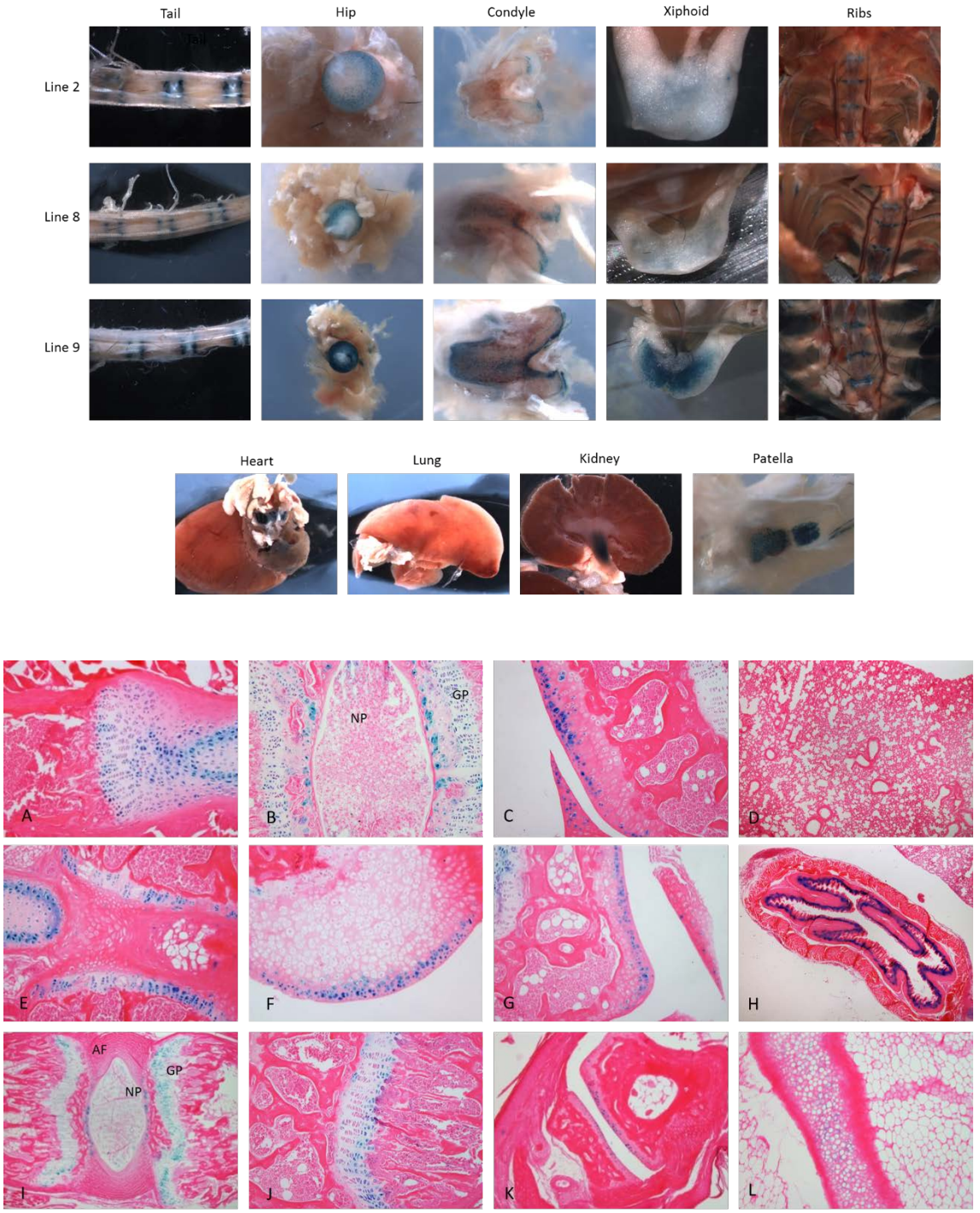
Examination of the -35kb lines at E12.5 shows minor differences between lines 2 (A), 8 (B) and 9 (C) but all share similarities in the limb staining (G-I) in whole mount. Histologically there is expression in the early limbs (L and O), ribs (K) and Meckel's cartilage (J) but no staining in the heart or lung (N) and little in the brain (M).

Figure 26: Whole mount images and histological examination of the -35kb at E13.5



E13.5 embryos expression whole mount from lines 2 (A) and 8 of the -35kb enhancer (B) showing skeletal elements with β-gal activity. Coronal view shows craniofacial staining but not complete ribs (C) and in the spine (E). Higher magnification of the vertebra (F) and the limbs (D). Histologically all chondrocytes are stained blue in Meckel's cartilage (A), ribs (B), skull (C and F), limb (D) and paw (E).

Figure 27: 8 week old mice lines of the -35kb enhancer whole mount and histology



[TOP] Whole mount images of 8 week old tissue taken from the -35kb lines 2, 8 and 9 shows similarities in staining for enhancer activity in the tail, hip, condyle and ribs with line 9 being the strongest at this age. There are weak staining in the xiphoid in lines 2

and 8 but stronger in line 9 which may be a copy number effect. There are other regions that the enhancer is able to express in the cartilage rings in the aorta and patella but not in the lungs or kidney.

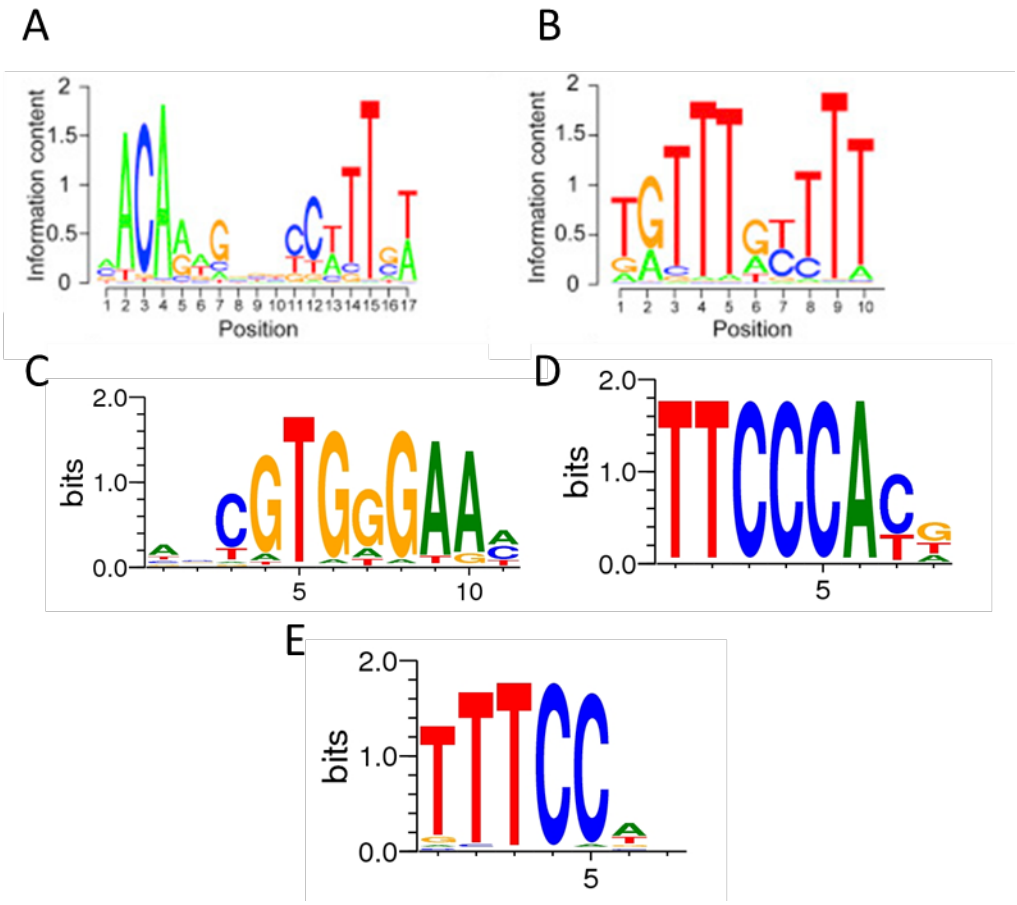
[BOTTOM] The -35kb expresses in 8 week old mouse in chondrocytes in the ribs expressing in the costal cartilage but stops at the ossified regions (A), growth plates [GP] from the sternum- costal cartilage (E) and knee (J); the superficial regions of the articular cartilage of the metatarsals (K), hip (F), condyle (C) and plateau of the knee (G) as well as the meniscus (C). In the tail and spine the enhancer is active in the end-plates and weakly in the annulus fibrosus [AF] but is absent from the nucleus pulposus [NP] (B and I). The enhancer does not express in the lungs (D) and very weakly in the xiphoid (L). The internal connective tissue of the aorta is marked by the -35kb (H).

4.1.4. Identification of transcription factor binding in the +26kb and -35kb enhancer

Enhancers can interact with multiple TFs at any one time, but key regulators have temporal importance and a loss results in loss of function on a transcription level. To identify and understand the TFs that modulate the activity of the +26kb and -35kb enhancers we used the program Biobase to screen for binding motifs and combined this with conservation by alignment using ClustalW to determine if each region was worth perusing (highly evolutionary conserved regions are more likely to be of significance). Sox9 has been previously been shown to activate the *Col2a1 (Bridgewater et al., 1998, Bridgewater et al., 2003, Lefebvre et al., 1997, Zhou et al., 1998)* and the *Acan* promoter and known enhancer (Sekiya et al., 2000, Han and Lefebvre, 2008, Doege et al., 2002) and it is classed as the master transcription factor that regulates the chondrocyte phenotype in development (Akiyama et al., 2002). We find that Sox9 is predicted to interact with both enhancers with strong degrees of conservation at 6 sites in the +26kb and 8 sites in the -35kb (Figure 29 and Figure 30 Red). Each region contain the single binding site of Sox9 or the double binding site in opposing direction with 3-4bp of separation seen in chondrocyte specific interactions (Bridgewater et al., 2003, Ohba et

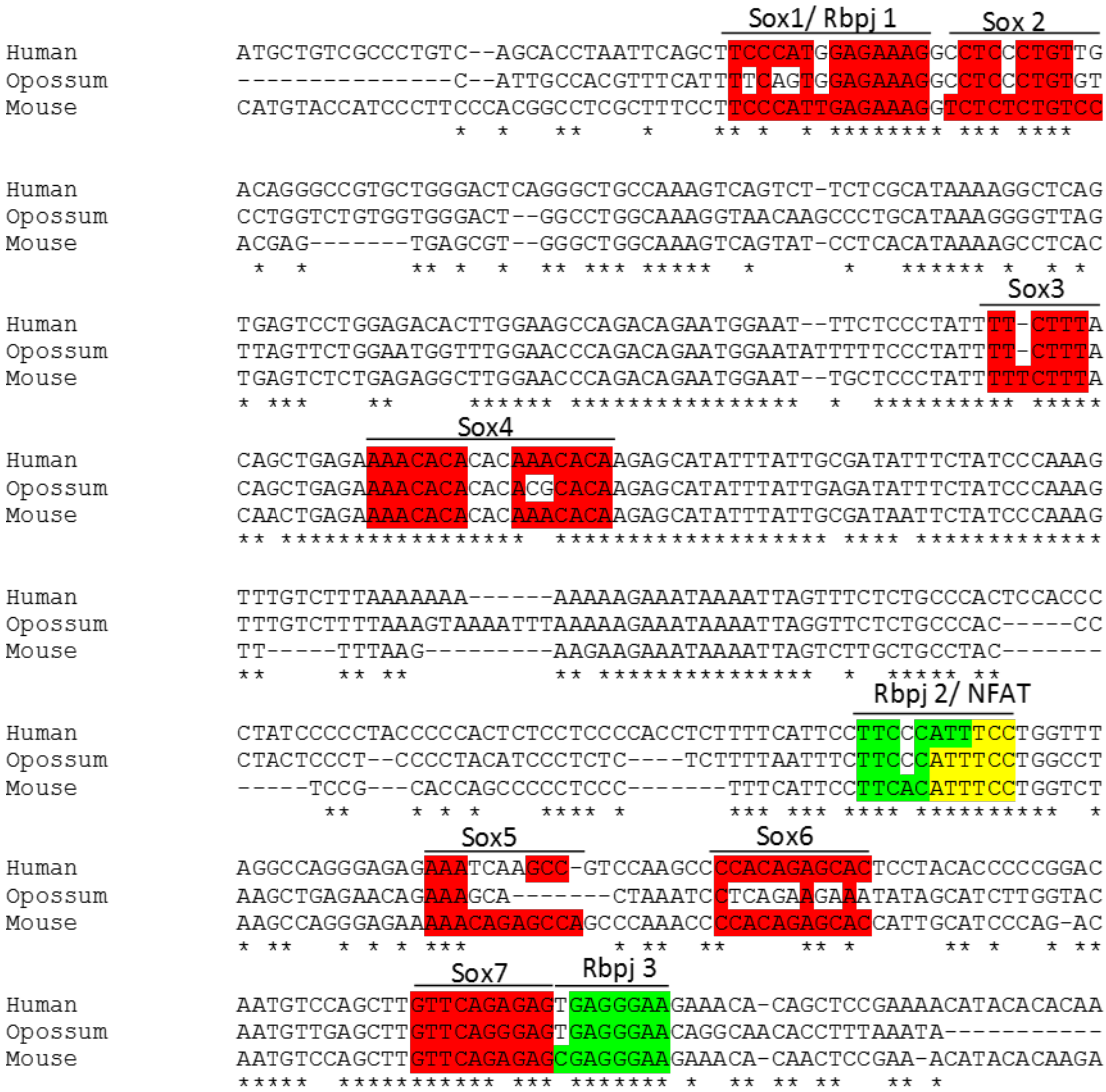
al., 2015). Rbpj-k is an important Notch signalling mediator and is important in maintenance of post-natal articular cartilage with a chondrocyte specific knock out using the *Acan*Cre(ERT2) postnatally resulted in OA like pathology (Liu et al., 2015, Mirando et al., 2013, Liu et al., 2016). Therefore we added Rbpj-k to the TF scree and found three conserved motifs in the +26kb and -35kb enhancer (Figure 29 and Figure 30 Green). NFAT1 was another TF taken into account due to the its age-dependant activity in chondrocytes, its regulation of Notch dependant differentiation and its role in regulating Sox9 in development (Lin et al., 2014, Zanotti and Canalis, 2013, Rodova et al., 2011, Wang et al., 2009). One binding site in each enhancer was identified for NFAT1 (Figure 29 and Figure 30 Yellow). The consensus sequence for each TF is shown in Figure 28.

Figure 28: Binding motifs of Sox9, Nfat and Rbpj-k



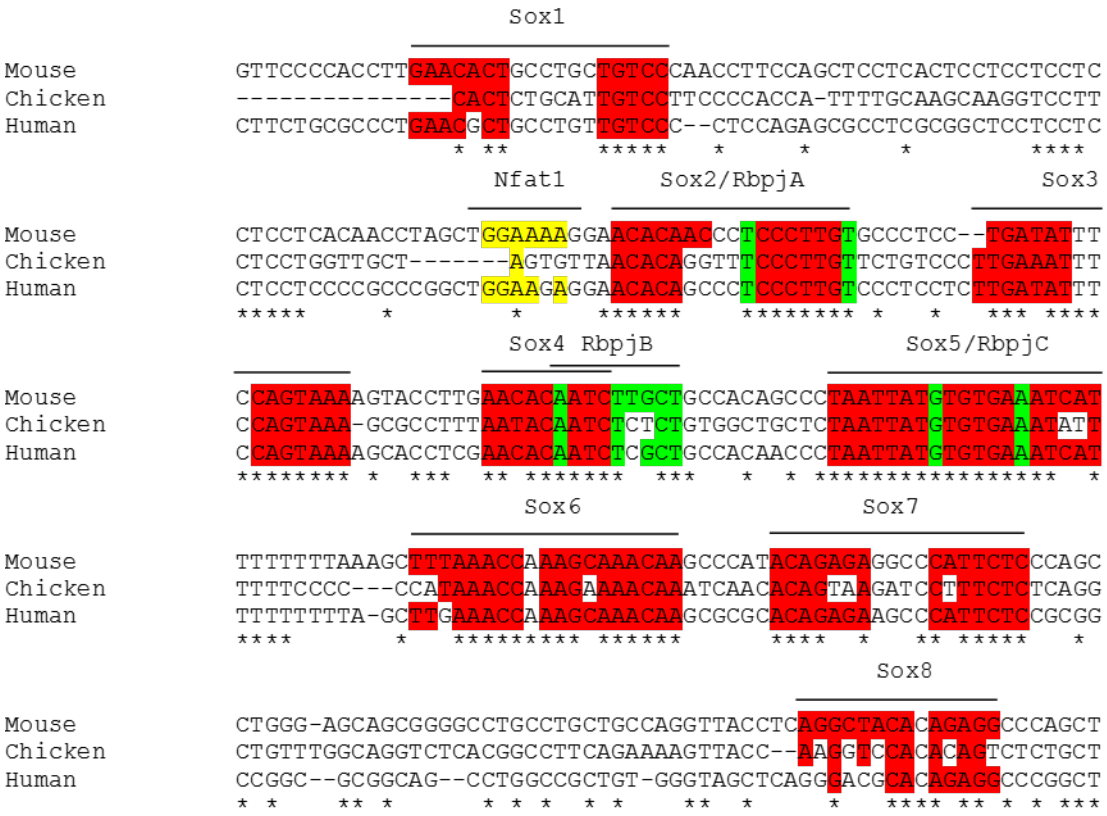
Binding sequences for (A-B) Sox9, the class 2 binding of a doublet Sox9 binding sequence reading in opposing directions separated by 304 nucleotides (A). (B) monomer sequence of Sox9 not generally associated with chondrocytes. (C-D) rbpj consensus binding sequence. (E) Nfat1 consensus binding sequence.

Figure 29: ClustalW alignment for the +26kb enhancer and possible transcription factoring binding sites



ClustalW alignment of the +26kb sequence between mouse, human and opossum. Using biobase seven Sox9 binding sites were identified (in red), three rbpj-k (green) and NFAT (yellow) that were examined using eletromobility shift assays.

Figure 30: ClustalW alignment of the -35kb enhancer and outline of possible transcription factor binding sites



ClustalW aligned mouse -35kb sequence to chicken and human with eight Sox9 binding sites (red), rbpj-k (green) and NFAT (yellow) that were examined using electromobility shift assays.

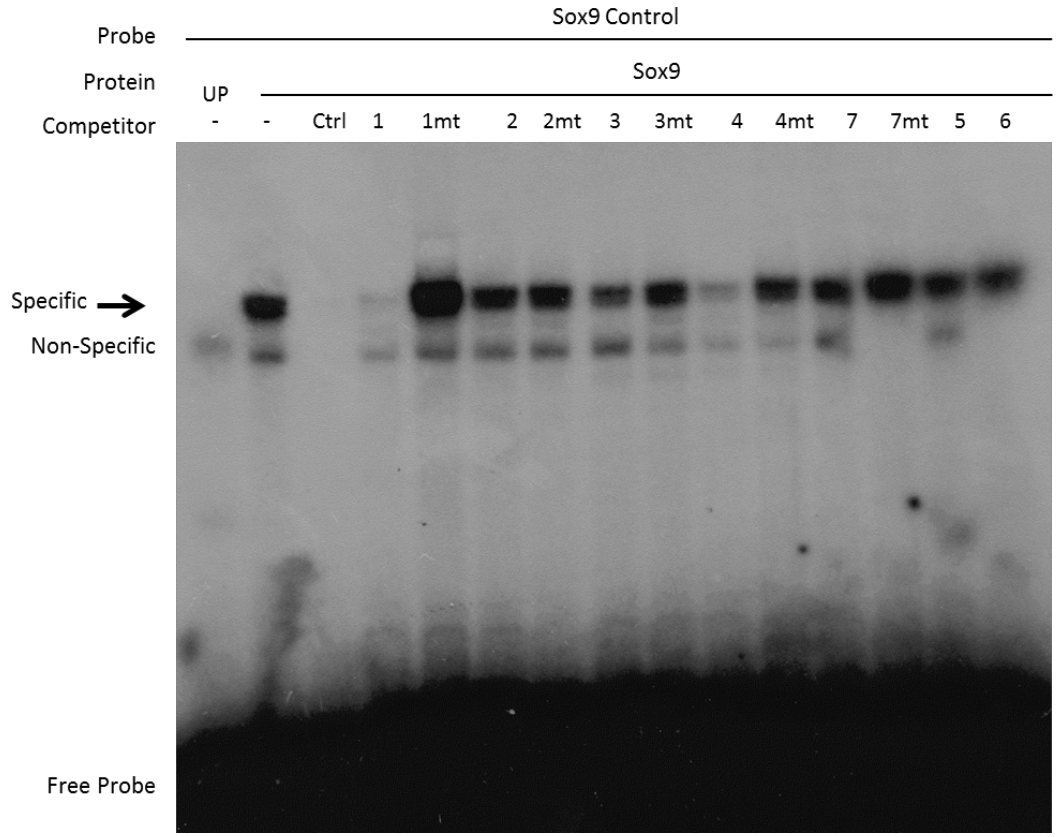
EMSAs provide a relatively quick method for detecting *in vitro* if a DNA region binds to the TF, as DNA bound to proteins migrating through a gel matrix moves slower than DNA alone you can determine, by labelling, the DNA or protein if a binding occurs by visualising the gel. The use of *in vitro* translated proteins from vectors to generate the proteins used in EMSAs removes the anonymity of the binding protein from nuclear protein extract and the requirement of using an antibody that targets the suspected protein (super-shift) which gives cleaner and clear binding patterns on a EMSA gel and more confidence in the identity of protein which has bound to the sequence. There are two types of competitive EMSAs, direct and indirect, the direct method labels each binding region by end labelling with P32, fluorophores or biotin and examines for binding for an appearance of a band on the gel. Indirect is the method utilised in this study due to the cost-effective and rapid screen of large amounts of binding motifs, this method labels a control or known binding motif for the TF. If the sequence that is predicted to bind is incubated with the protein first it should bind reducing the free protein for the labelled control sequence to bind to, this mean the band is diminished if the sequence of interest is interacting. To ensure it is the binding site that was predicted is the site that is binding we introduce point mutations into the oligo (mt) and this should reduce the efficiency or completely abolish the binding interactions, this ensure that in designing the oligos there was no accidental introduction of binding sites.

Sox9 interacted with the +26kb enhancer at sites 1 and 4, albeit weakly in both cases, where mutations in the binding sites resulted in the control oligo binding (Figure 31). Initially, the -35k sites were interrogated using increasing concentration of the oligos without mutations, which identified sites 2 and 3 with good competition and site 7 as another site (Figure 32). Site 4 on the 100% concentration lacks a band due to technical error. Each site was then examined alongside their mutated version showing that mutations into the Sox9 binding motif removed the binding affinity and reconfirmed the binding ability of the three sites (Figure 33).

The EMSA for rbpj-k reveal that the protein has two interaction sites in the +26kb, site 1 and 3 however the mutations in site 1 were unable to adequately block the binding

(Figure 34A). In the -35kb only one site was able to bind, Site C, but we were unable to produce a clean EMSA gel to demonstrate the ability of the mutations to diminish binding (Figure 34B). In terms of NFAT1 binding both enhancers were able to bind to the NFAT core binding domain protein (Figure 35).

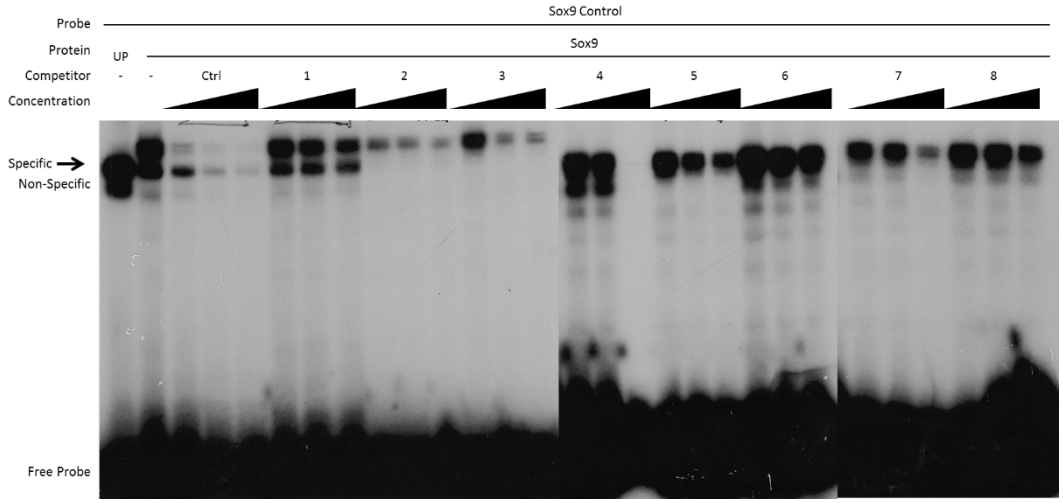
Figure 31: Sox9 indirect competitive EMSA for the +26kb enhancer



Indirect competitive electomobility shift assay for the 6 regions of the +26kb possible Sox9 binding using a labelled Sox9 control probe. Lane 1 UP is the lysate from the in vitro translation reaction with no vector to determine any non-specific binding. The second lane is the probe with no competitor to show it is binding to the proteins. The third lane is the control competitor therefore the band should be removed if it is able to bind. Sites

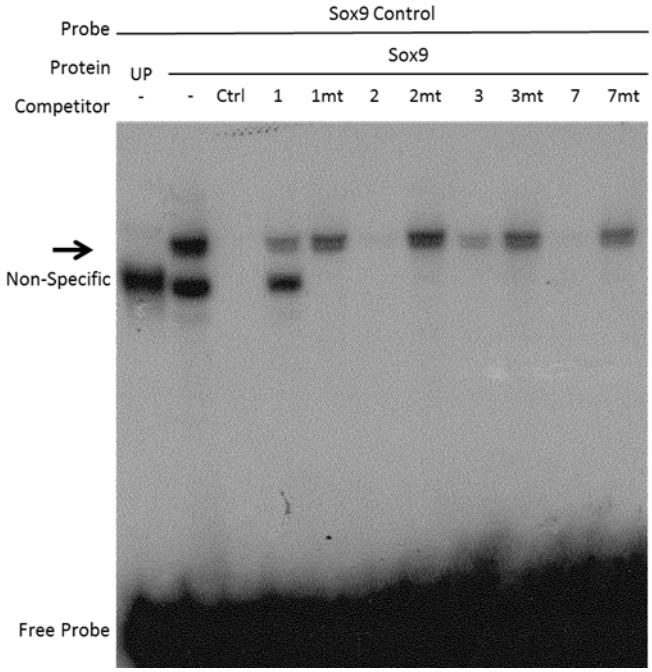
1 and 4 are able to bind the Sox9 protein and introduction of mutations into the binding sites (1mt and 4mt) diminishes the binding activity so the band reappears.

Figure 32: Sox9 indirect competitive EMSA for the -35kb enhancer with increasing concentration of competitor



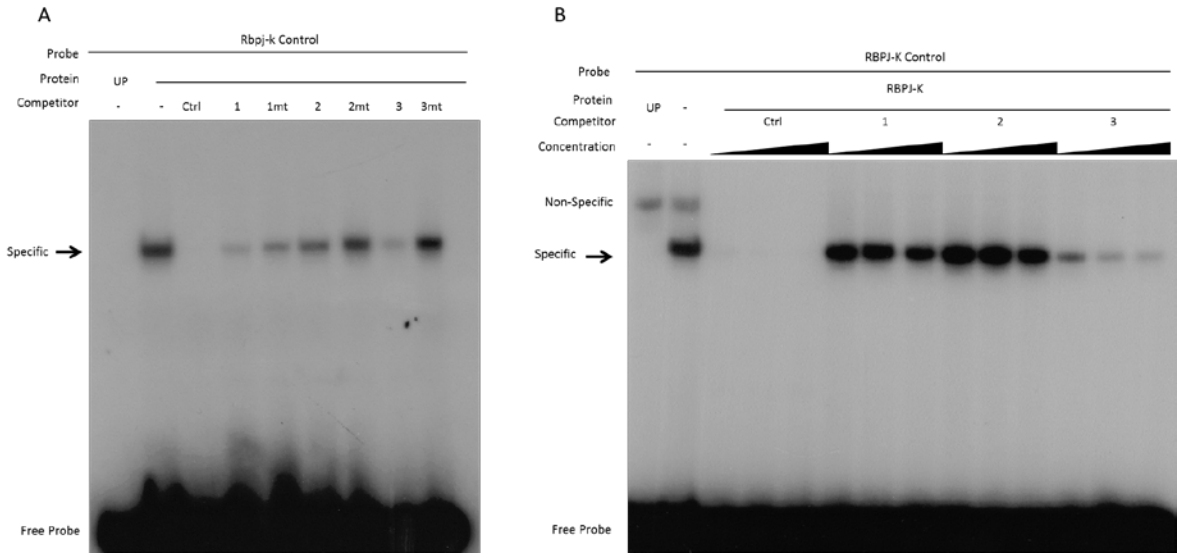
Indirect competitive electromobility shift assay using a Sox9 control probe that is able to bind to Sox9 protein for the -35kb enhancer region. Lane 1 UP is the lysate from the in vitro translation reaction with no vector to determine any non-specific binding. The second lane is the probe with no competitor to show it is binding to the proteins. The third land is unlabelled control oligo (competitor) it is able to compete and bind to the protein so the signal is diminished when increasing the concentration of the control oligo. Each competitor is increased by 20%, 50% and 100%. Sites 2 and 3 are able to compete with the control probe to remove the band therefore binds to Sox9 in vitro. Site 7 only with the highest concentration is able to compete with the control probe. Site 4 appears to bind but this was done in error, the probe was not added into the reaction.

Figure 33: Sox9 indirect competitive EMSA for the -35kb enhancer with mutated sequence



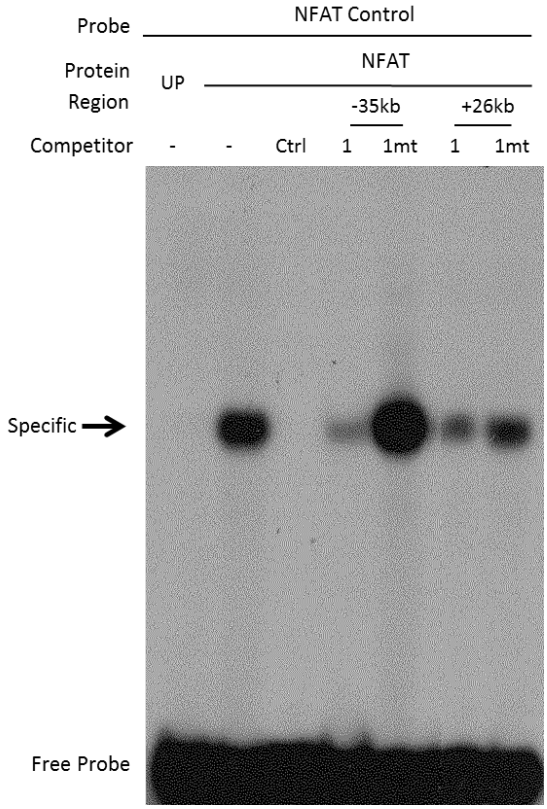
Indirect competitive shift assay. Lane 1 UP is the lysate from the in vitro translation reaction with no vector to determine any non-specific binding. The second lane is the probe with no competitor to show it is binding to the proteins. The third lane is the control competitor therefore the band should be removed if it is able to bind. This confirms that sites 2, 3 and 7 of the -35kb is able to bind to Sox9 in vitro but not site 1. Introduction of mutations into each binding sites removes the binding to Sox9 (2mt, 3mt and 7mt).

Figure 34: Rbpj-k indirect EMSA for the +26kb and -35kb enhancer



Indirect competitive shift assay for Rbpj-k for enhancers +26kb (A) and -35kb (B). Lane 1 UP is the lysate from the in vitro translation reaction with no vector to determine any non-specific binding. The second lane is the probe with no competitor to show it is binding to the proteins. The third lane is the control competitor therefore the band should be removed if it is able to bind. Site 3 is able to bind to Rbpj-k in vitro and mutations introduced to the Rbpj-k binding site abolishes the binding. Site 1 was able to interact but mutations could not remove the binding as effectively. For the -35kb Rbpj-k there was an increase of competitor from 20%, 50% to 100%, only site 3 binds to Rbpj-k protein in vitro.

Figure 35: The -35kb and +26kb enhancer binds with NFAT core binding domain with different affinities

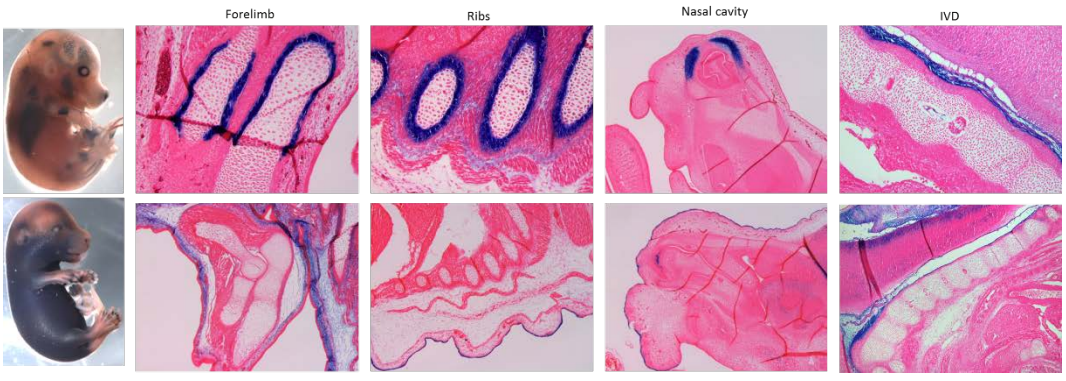


Indirect competitive Electromobility shift assay using the NFAT core protein binding to detect interactions within the +26kb and -35kb enhancer. The NFAT control oligo was used as the probe and an unlabelled oligo of the same sequence is the control competitor. Lane 1 UP is the lysate from the in vitro translation reaction with no vector to determine any non-specific binding. The second lane is the probe with no competitor to show it is binding to the proteins. The third lane is the control competitor therefore the band should be removed if it is able to bind. Both NFAT sites in the +26 and -35kb are able to bind weakly to the NFAT protein and introduction of mutations into the sequences diminishes the binding.

4.1.5. *In vivo* analysis of transcription factor binding site mutations

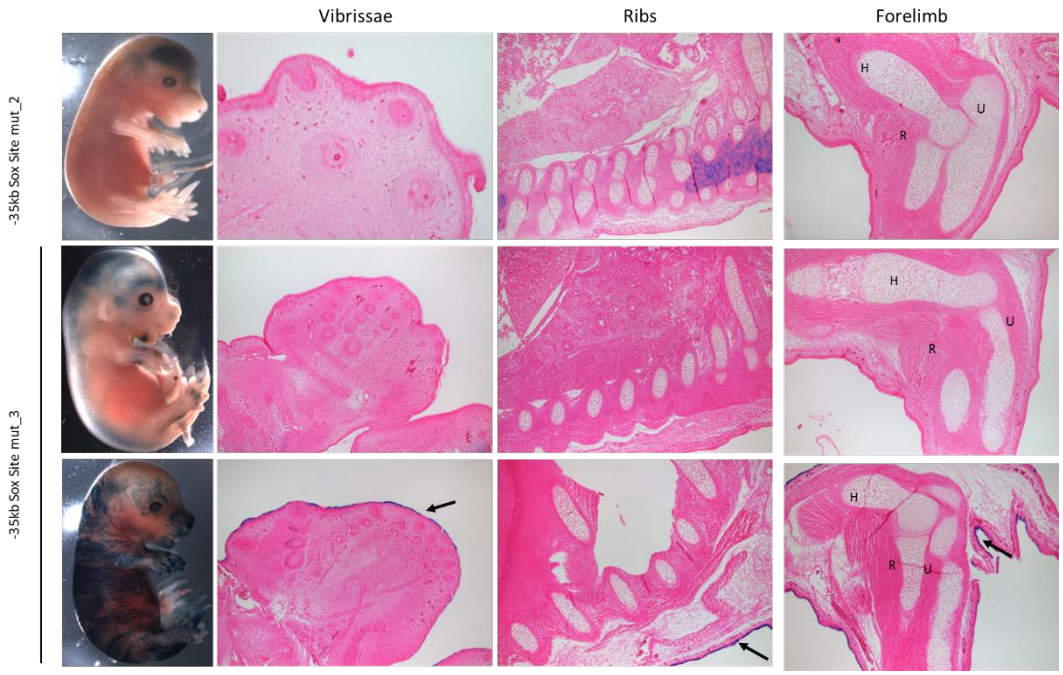
To assess the importance of the TF binding site we cloned the mutated sequences into the HSP68/GW cassette and generated transgenic mice, to assess the role of Sox9 in the -35kb we introduced mutations into Sox9 binding sites 2, 3 and 7 together and collected at E15.5, the same time point we observe staining in the WT. We expected for Sox9 mutations to completely abolish the expression of the enhancers. Therefore, to our surprise the transgene was still active but no longer in the chondrocytes, the expression had shifted into the fibroblast of the skin and in some cases the perichondrium there were no longer expression in the chondrocytes, in the ribs, limb and IVD (Figure 36). To determine if a single site was important and responsible for the shift in expression we injected each site mutated individually. Sox9 sites 2 and 3 showed no chondrocyte specific expression in the limbs or other cartilage based tissue (Figure 37). Site 3 did present with superficial staining although this requires more transgenic animals to be generated to confirm. Site 7 looked as though there were no chondrocyte specific expression when observed whole mount (Figure 38A), with more staining in the vibrissae. However, there appears to be a population of chondrocytes that are active with the transgene seen in the forelimbs radius and ulna (Figure 38F and G), but no staining is observed in the metacarpals, metatarsals, phalanges (Figure 38H and J), ribs or ear. By eye there seems to be no defined characteristics for the chondrocytes that are expressing the transgene as the sagittal section of humerus (Figure 38E) show no chondrocyte staining in the proliferating, pre-hypertrophic or hypertrophic cells.

Figure 36: The -35kb Sox mut 2_3_7 at E15.5 abolishes expression in the chondrocyte



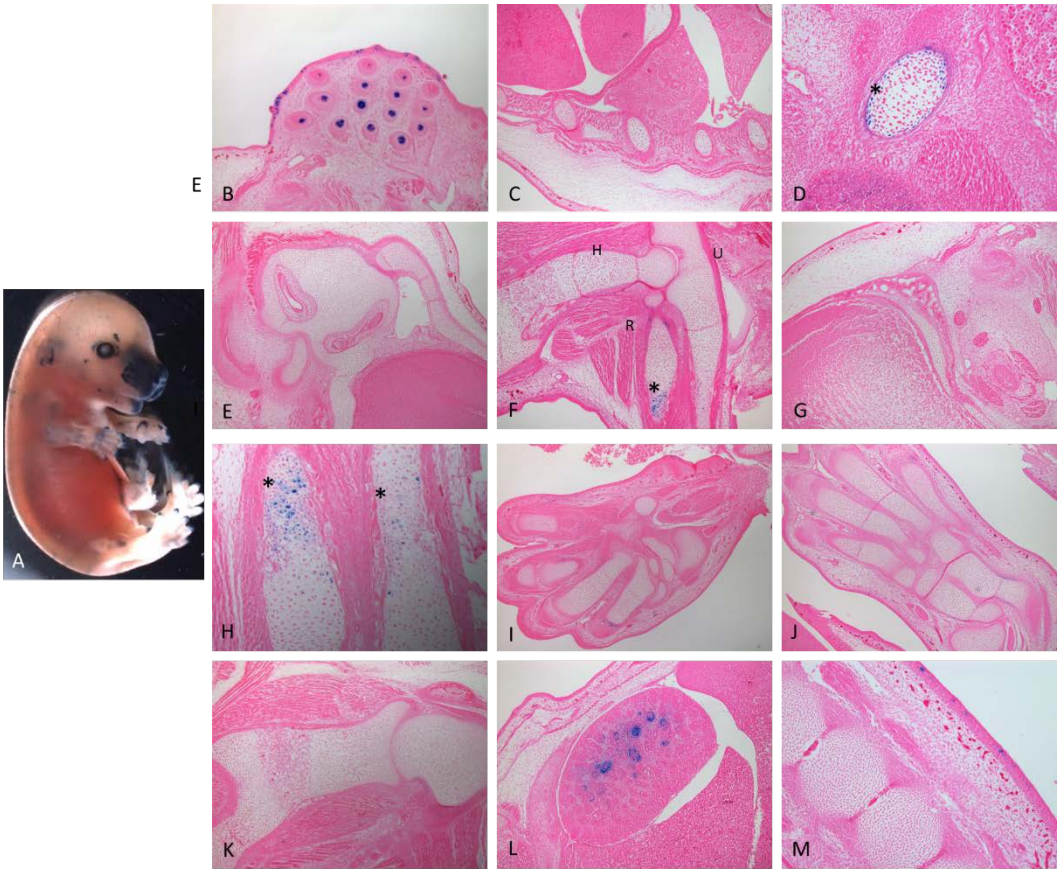
Mutations in all three of the sox9 binding sites in the -35kb enhancer abolishes the expression in the chondrocytes. Two different expression is seen which may be as a result of SOI effects. The first (Top) still remains in skeletal elements but the perichondrium seen in the forelimb and ribs. No expression is seen in the vertebra and IVD but some in the meninges. The second expression patter resides in the epidermis or developing dermis of the skin (Bottom).

Figure 37: The -35kb Sox mut_2 and Sox mut_3 at E15.5 are no able to express in the chondrocyte



Individual mutations in the Sox9 binding sites of the -35kb enhancer removes the expression in the chondrocytes and other tissues. Binding sites 2 alone (top) only presented with notochord expression. Binding site 3 (Bottom) also abolishes the expression of the enhancer albeit there are some ectopic epidermal expression. In both cases the previously observed expression of all three sites mutated is not reproduced.

Figure 38: The -35kb Sox mut 7 expresses weakly in the chondrocyte

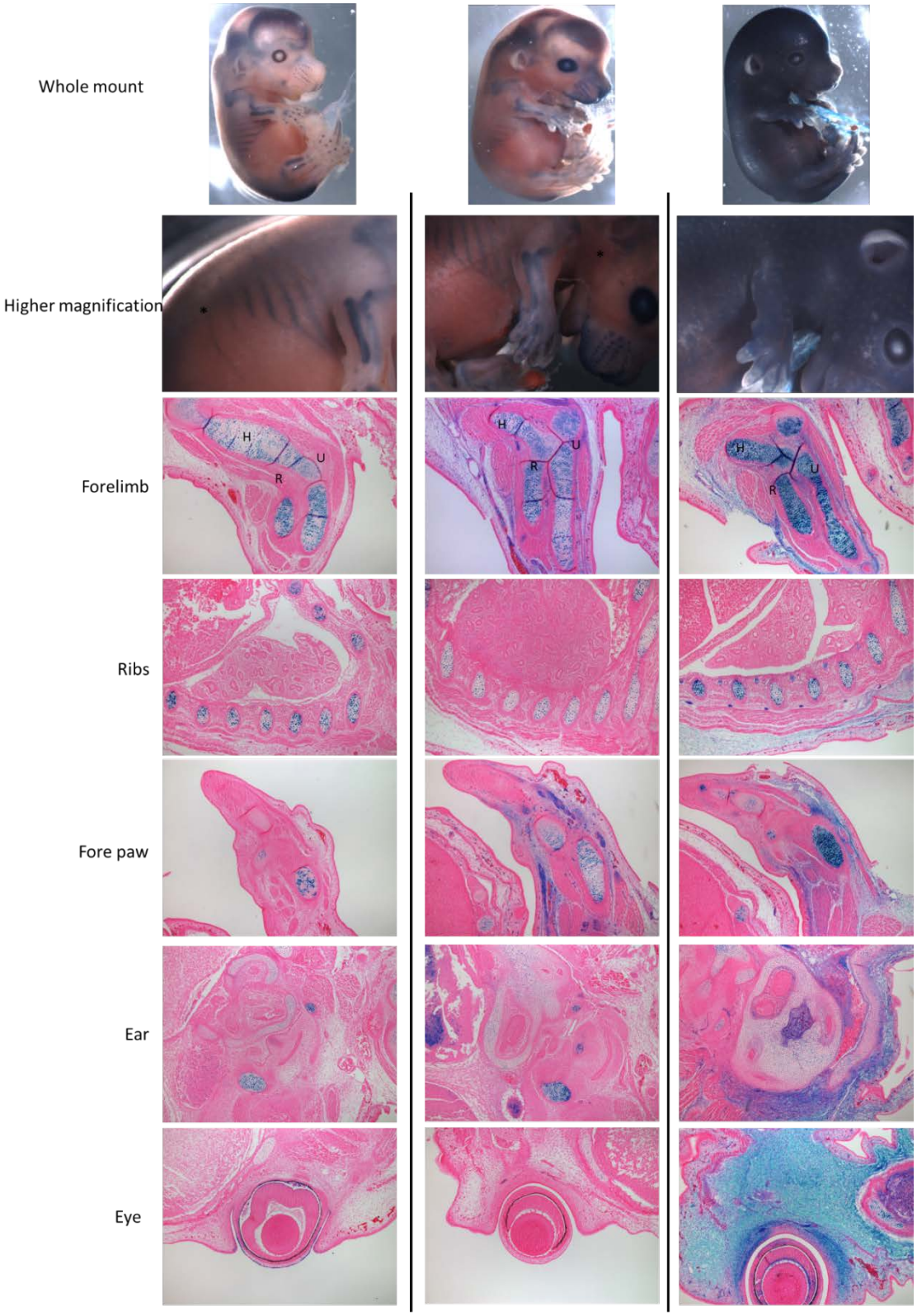


The Sox9 binding site 7 of the -35kb enhancer when mutated removes the expression in almost every chondrocyte. However there is expression in the forelimb central chondrocytes in the radius [R] and ulna [U] but not in the humerus [H] (F and higher magnifications G). There is expression in the vibrissae (B) and in the glomeruli of the kidney (I). (A) is the whole mount of a representative embryo; (C) ribs, (D) ear; (E) humerus; (H) front paws and rear paws (J).*

We generated -35kb rbpj-k mutated sequence transgenic mice to interrogate the effect. On visual inspection there appears to be no effects on the majority of the embryos at E15.5 but some did stain (Figure 39). At E15.5 it appears that the rbpj-k binding site has no effect on the chondrocyte specificity, although the staining is more scattered. There

is increased ectopic expression in the fibroblast in the skin (Figure 39) and it is noted that there is increased vasculature expression. Table 8 gives an overview of how many constructs were expressed and Table 9 summarises the expression of each of the TF mutations tested.

Figure 39: Rbpj-k binding site mutations at E15.5 does not affect chondrocyte expression



Mutations in the Rbpj-k binding site in the -35kb has little effect on the chondrocyte specific expression of the enhancer. However, the expression appears more scattered and more expression in the skin in 3 of the 4 embryos that expressed. [H] humerus; [R] radius; [U] ulna. There appears to be vasculature expression not seen in the wildtype enhancer [*].

Table 8: Overview of expression verses number of expressing of transcription factor binding sites mutations at E15.5

Construct	**Expresser/ Transgenic (%)**	**Total embryos analysed in transient transfection**
-35kb WT	4/4 (100)	17
-35kb Sox mut_2_3_7	3/3 (100)	30
-35kb Sox mut_2	1/2 (50)	15
-35kb Sox mut_3	4/4 (100)	14
-35kb Sox mut_7	2/4 (50)	9
-35kb rbpj_mut_C	4/5 (80)	20

"Expresser" are the embryos which presented with β-gal expression. "Transgenic" were the embryos genotyped positive for LacZ and the percentage (in brackets) is the number of genotyped positive embryos compared to the expressers. The final column is the total embryos analysed in a litter.

Table 9: Summery of binding site mutation expression

Tissue	Intensity of staining				
	-35kb Sox mut 2_3_7	-35kb Sox mut 2	-35kb Sox mut 3	-35kb Sox mut 7	-35kb Rbpj mut_ C
Chondrocranium	-	-	-	-	++S
Nasal cartilage/ olfactory	±	-	-	-	-
Meckle's cartilage/ mandible	-	-	-	-	++S
Costal cartilage of the ribs	-	-	-	+	+++S
Femur	-	-	-	-	+++S
Tibia	-	-	-	-	+++S
Fibula	-	-	-	-	+++S
Digits of hindlimb	-	-	-	+	+++S
Digits of forelimb	-	-	-	+	+++S
Scapula/ humerus	-	-	-	-	+++S
Ulna	-	-	-	+	+++S
Radius		-	-	+	+++S
Intervertebral discs	-	-	-	-	+S
Lumbar spine	-	-	-	-	+S
Cartilage of the ear/ occipital	-	-	-	-	+S
Tail	-	-	-	-	++S
Brain	-	++	±	-	-
Vibrissae	-	-	-	++++	++++
Heart	-	-	-	-	-
Lung	-	-	-	-	-

Liver	-	-	-	-	-
Tongue	-	-	-	-	-
Skin	±±±	-	±±	-	±±±±
Kidney	-	-	-	+++	-
Blood vessels	-	-	-	-	±±

Summary of the staining intensities and location for the +26kb and -35kb enhancer. – is not detected; + weakly detected; ++ not completely uniform staining throughout tissue; +++ majority of tissue stained; ++++ strong staining in entire tissue; ± staining present in some founder embryos but not all. "S" dictates that the expression was not uniform but scattered or "patchy" in some areas of the tissue.

4.2. Two Acan cis-elements that have defined developmental roles

Many genes can contain several DNA regulatory elements, the most characterised being the β-globin enhancer (Caterina et al., 1994). Each element may be responsible for temporal and spatial activation of a gene, guiding the precise activation for development and normal homeostasis and when it is no longer needed, it eventually become inactive. Inactive enhancers have also been shown to be essentially in long range activation of genes, for example in B cells the *Igk* enhancer MiEk is inactive but interactions with the active enhancer *Tcrb* enhancer Eβ changes the localisation of CBFβ controlling *Tcrb* recombination (Proudhon et al., 2016). Currently, it is rare to find cartilage proteins that have greater than two enhancer elements to varying degrees. For example, the *Col10a1* is under influence of tissue specific enhancers located between -2410 and -1875bp (Gebhard et al., 2004, Beier et al., 1997) and -4.4 to -3.8kb (Zheng et al., 2009) that together follow the complete expression of *Col10a1* that are intergenic. The *Col11a2* has two chondrocyte specific enhancers (Bridgewater et al., 1998), the *Col27a1* contains at least three enhancers in the first intron (Jenkins et al., 2005). However, the *Col2a1* is regulated by intronic sequences, an 18bp region in the first intron (Rossi and de Crombrugghe, 1987, Leung et al., 1998, Zhou et al., 1998, Lefebvre

et al., 1996), the 7th intron (Shinomura et al., 2012), in the 2nd intron that regulates splicing (McAlinden et al., 2005). Col1a2 on the intergenic enhancers regulates fibroblasts at -19.5kb (Bou-Gharios et al., 1996, Antoniv et al., 2001, De Val et al., 2002). So we wanted to determine if *Acan* is under a higher order level of transcription, seeing as there is a large region of non-coding similar to *Sox9* which houses multiple enhancers that drives different expression in different tissues (Mead et al., 2013).

4.2.1. The -65kb enhancer element of Acan is able to express in chondrocytes at E15.5

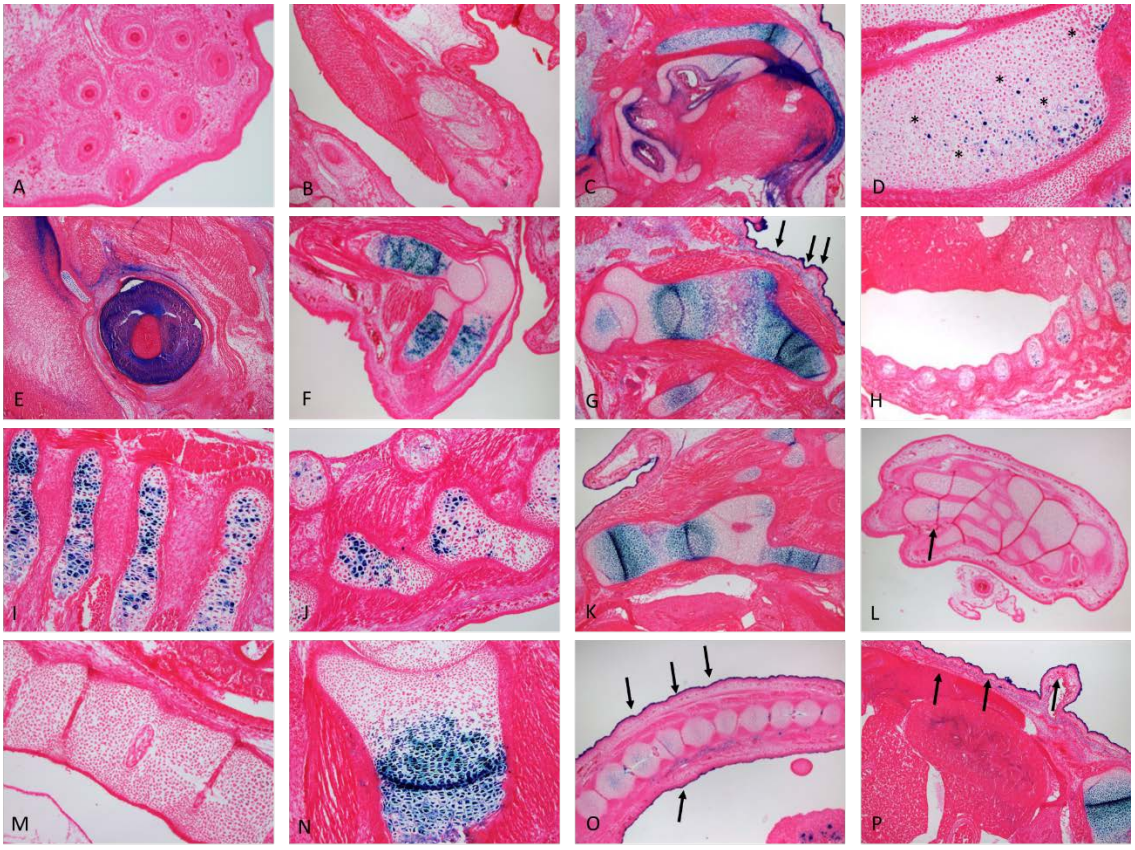
After identifying the -65kb region upstream of the *Acan* TSS as a possible enhancer, the region was cloned into the HSP68 gateway vector and analysed transiently to determine if it was able to express. At E15.5 there were two expression patterns seen in whole mount (Figure 40A -C). Skeletal element staining were observed in the core regions of the forelimbs, hindlimbs and in the ribs (Figure 40D and F). However, staining was not distinguished in the paws or the tails in one embryo. In the second there was mass staining in the skin making it difficult to distinguish if there were skeletal staining (Figure 40C, E and G). Histologically, the -65kb at E15.5 is expressed by the chondrocytes but unlike the +26kb and -35kb it is not in every chondrocyte. In the skull there is strong staining in the ventral bone but closer magnification of the dorsal regions shows only a scattered few cells are staining for B-gal activity (Figure 41C and D). The limbs present a staining pattern that is reserved for the core regions of the long bones of the radius, ulna and humerus (Figure 41F and G), and femur (Figure 41K), closer inspection of the histology sections shows that the cells appear to be mainly hypertrophic chondrocytes of the radius, ulna and humerus (Figure 41N) No staining is observed in any regions of the IVD (Figure 41 M), Meckel's cartilage (Figure 41B), vibrissae (Figure 41A), in the paws (Figure 41 L), tail (Figure 41 O) or the kidneys (Figure 41 P). Scattered strong staining is seen in the vertebral body (Figure 41J) and ribs (Figure 41H and I). Staining in the superficial skin, near the epidermis is seen in the embryos that expressed in the skin (Figure 40C and Figure 41K and O). Lastly, the eye appears to express and stains with X-gal in the vitreous humor (Figure 41 E).

Figure 40: Whole mount images of the -65kb enhancer

Whole mount images of the two expression patterns seen in the -65kb enhancer at E15.5. (A) Sagittal image of a representative embryos with core element staining in the limbs and ribs. (B) rear view showing vertebrate staining. (C) other staining pattern seen in whole embryos of complete skin staining (D) a higher magnification of the forelimb

and ribs displaying staining and the hindlimb (F). (E) shows skin staining in the limbs and in the vibrissae, however craniofacial features are difficult to decipher (G).

Figure 41: Histological examination of the -65kb enhancer



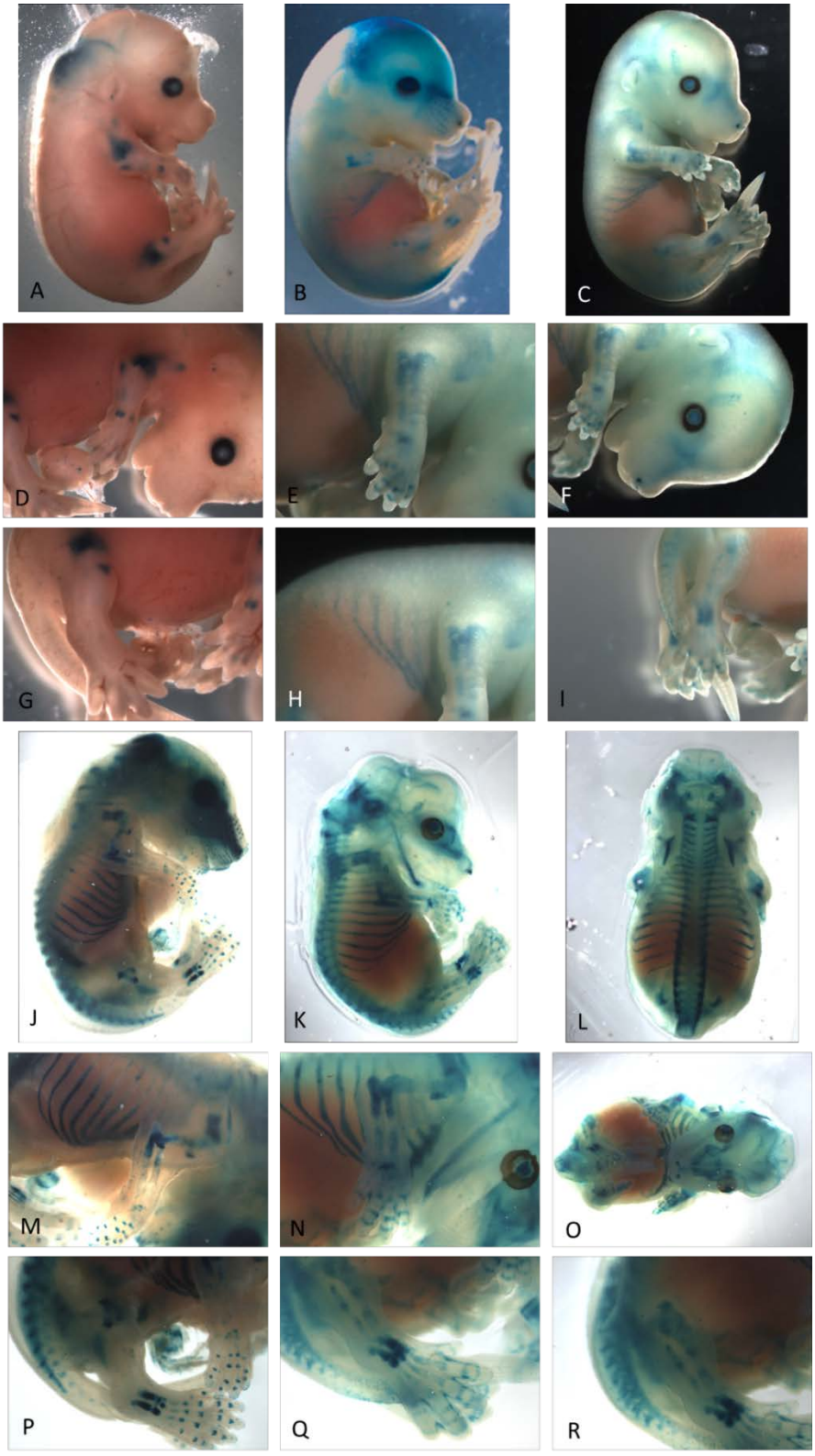
Histological examination of two -65kb enhancer at E15.5. C, G, K, O and P come from the pan skin expressing embryo with t. Staining is seen in chondrocytes but of a select subset, in the forelimb (F) it localises in the hypertrophic chondrocytes. In the ribs, (H and I), vertebrae (J), tail (O) and scattered in the skull (*) (D) it appears scattered to specific chondrocytes and in the carpels of the fore paw there is some but not a significant amount (L). Staining is absent from the vibrissae (A), Meckel's cartilage (B) and IVD (M).The eye stained with X-gal (E) as did the cutaneous layer of the skin in some of the embryos identified by the arrows (O and P). The -65kb does not express in lungs (H) or in the kidney (P).

4.2.2. Characterisation of the far up stream enhancer -87kb in aggrecan gene

Transient transgenic mice at E15.5 harbouring the -87kb enhancer construct presented staining in skeletal elements (Figure 42), therefore a line was established for further analysis. Whole mount at E15.5 showed joint and ribs staining with X-gal but there was little or no staining in other areas (Figure 42H). On the limbs it appears confined to the central regions and in the paws it resides in the joints between the distal, proximal and metacarpals (Figure 42E and F). However, one line expressed differently to the others, in the craniofacial region, we cannot determine if this is real or if there were possible SOI or copy number effects where on some occasions there would be staining in areas where the transgene lands in an open chromatin (Figure 42B) and the limbs expressed more weakly in comparison (Figure 42A, D and G).

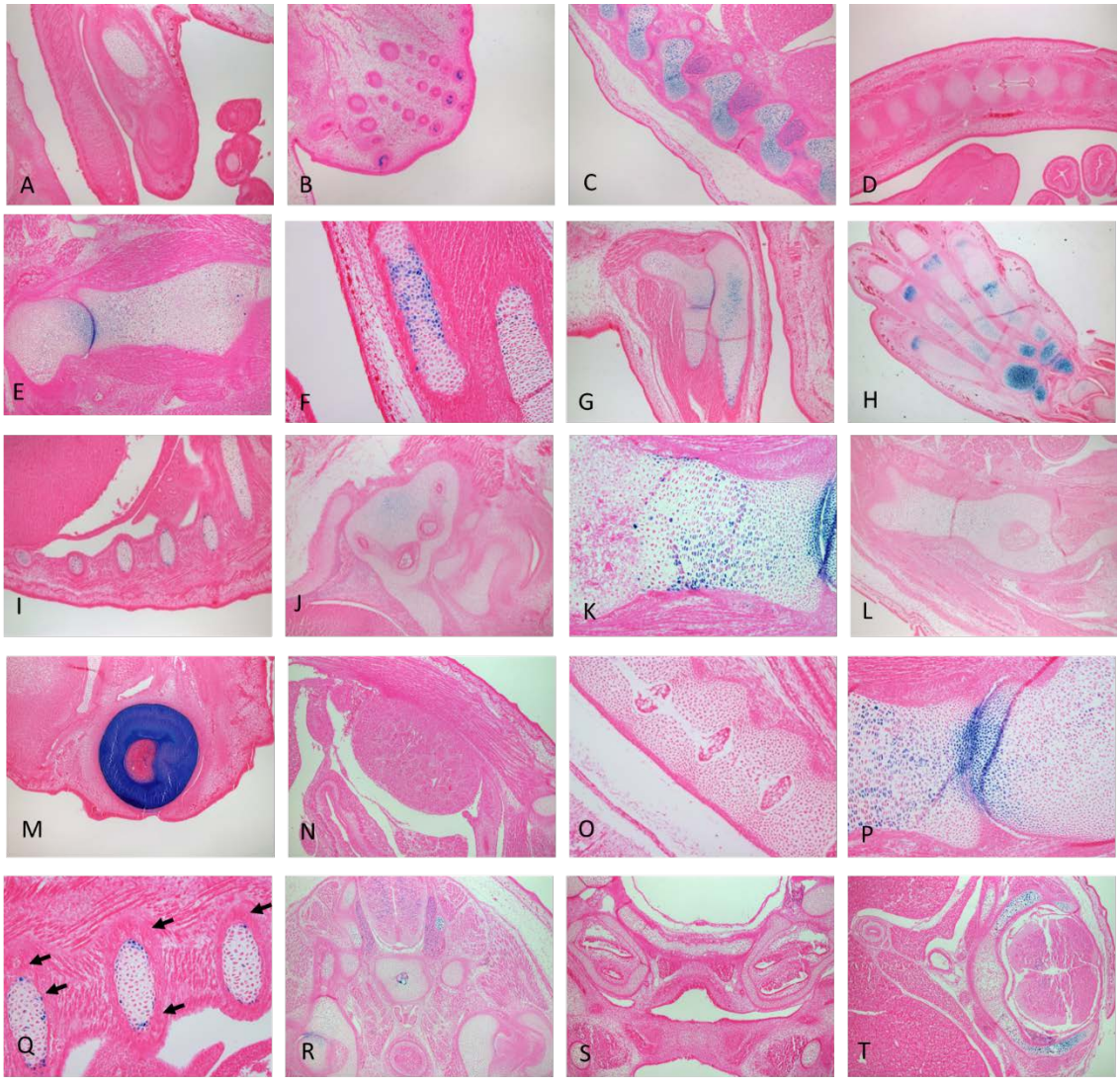
Histologically, the -87kb does not express in the IVD (Figure 43O and R), tail (Figure 43D), jaw (Figure 43S), kidney (Figure 43N) and in Meckel's cartilage (Figure 43A). There is weak expression in the ear chondrocytes (Figure 43J) and in the femur (Figure 43L) and more scattered and superficial staining seen in the forelimb (Figure 43G) and ribs (Figure 43I and Q). Closer inspection of the radius (Figure 43F) and humerus (Figure 43E, K and P) show that the chondrocytes that stain appear to move in the direction of the growing cartilage towards the articular cartilage. Additionally, this pattern of favouring the joints is seen in the entire limb (Figure 43G). There is strong consistent staining in the eye (Figure 43M) and of the chondrocytes of the vertebra seen in transverse (Figure 43T) coronal (Figure 43C) sections. In the paw the carpals are strongly stained whereas when the chondrocytes migrate towards the distal areas like the phalanges the staining is localised to the joints (Figure 43H).

Figure 42: Whole mount images of the -87kb enhancer



Whole mount images of expression patterns seen for the -87kb enhancer at E15.5. (A) Little expression seen mainly in the joint in the hips and shoulder, and patterns is noted in the paw and rear of the head. (D) Show higher magnification of the forelimb and (G) is a higher magnification of the hindlimb. (B and C) shows strong staining in the ribs (H) and in the bones of the fore and hind limb. There is intense staining between the bones of the paws (E and F). In the hindlimb (I) the staining is more localised into the core regions. Clearing the embryos with KOH shows greater detail in the skeletal elements showing that the -87kb expressing in the skeletal elements but not in the distal limbs where it resides in the joints (J-R).

Figure 43: Histology images of the -87kb enhancer



Eosin counter stained sections of a representative of the -87kb at E15.5. The staining is scattered marking certain chondrocytes in the humerus (E), ribs (I and Q), femur (L), ear (J), radius and ulna (G and F). The chondrocytes are patterning towards to ends of the bones and primordial articular cartilage seen in higher magnification of the humerus (K and P), the vertebra and the carpels are stained uniformly but the metacarpals are stained more towards the forming articular cartilage (H). The IVD (O), tail (D), kidney (N) and Meckel's (A) does not stain and only the proximal vibrissae express. Additionally,

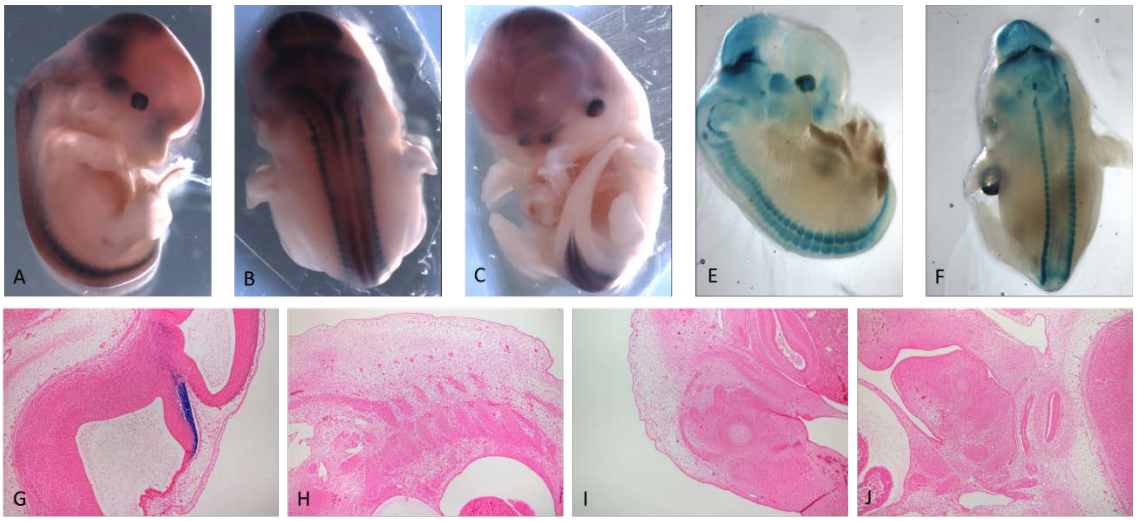
there is staining is seen in the eye and transverse sections show that there is little staining in the lumbar region (R) and jaw (S) until the costal vertebra (T).

4.2.3. Interrogation of the -87kb at different developmental ages

The -87kb enhancer favours expression in the joints. Therefore, the next step was to determine if that expression changes with age. At 8 weeks of age the -87kb expression is strong in the articular cartilage seen whole mount tissues of the knee (Figure 46G and H), femoral head (Figure 46B). In the hind paw the joints are stained strongly (Figure 46 C), as well as the growth plates of the ribs (Figure 46A) and cartilage end plates of the tail (Figure 46I). The patella (Figure 46E) and xiphoid (Figure 46D) but no staining is seen in the kidneys with the exception of some endogenous staining (Figure 46F). Histologically, the growth plate chondrocytes stain in both the sternum (Figure 46J) and the cartilage end plate, but in the IVD there is staining surrounding the nucleus pulposus in the annulus fibrosus, but not in the nucleus itself (Figure 46L). The articular cartilage chondrocytes are expressing in the superficial regions in the knee (Figure 46K and Q) and femoral head (R) cartilage. However it must be noted that the images presented as evidence the X-gal staining appears non-uniform, strong in some cases or non-consistent between different mice, although they are from the same line. One explanation for this is due to penetration issues of the X-gal stain this may have arisen from failure to open up the joint at the time of staining or tissue processing, this means the substrate or processing solutions were unable to penetrate into the central parts of the knee joints. We are certain that the inconsistencies are not a result of the transgene position effect or copy number as these mice are different from the same litter from one founder. Furthermore sectioning errors that rose from poor processing or poor technique that resulted in loss of samples or folding of the sections made it difficult to select images to present in the data set. The images are chosen as the best representation and the clearest image of chondrocyte staining in the knee. In the paw the articular cartilage chondrocytes are the only cells to express B-gal (Figure 46N) this is

seen more clearly in a higher magnification (Figure 46O). In the ribs it is mainly the superficial cells that are expressing (M) and in the xiphoid there is continuous staining throughout (Figure 46P).

Figure 44: E12.5 whole and histology images of the -87kb enhancer

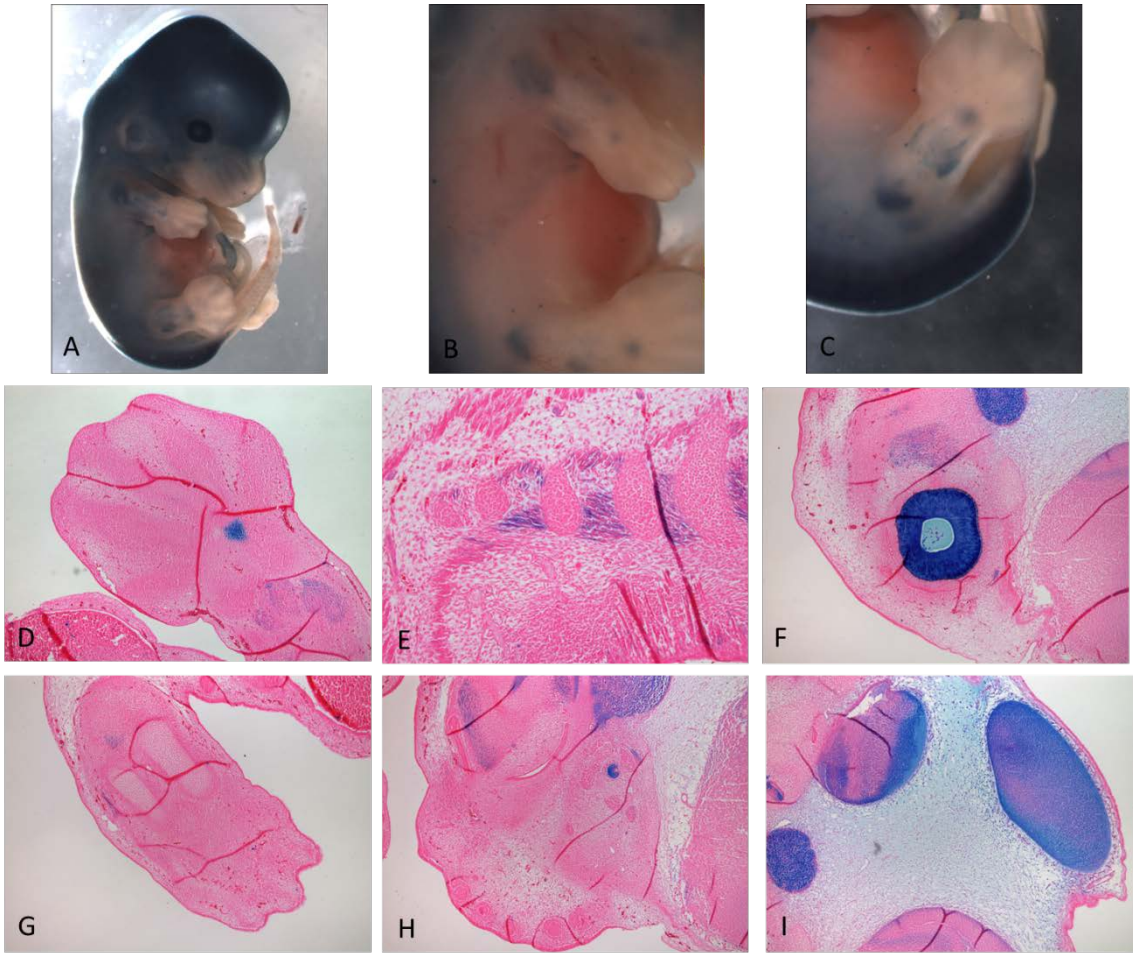


At E12.5 the -87kb is absent from the limbs seen in whole mount (A-C) and clearing the embryos with KOH yields similar observations (E and F) and by histological examination (I). Additionally, it is not expressed in the forming ribs or jaw (H and J). It is present in the brain (J) and although histological sections are absent it remains in the somites.

To determine the earliest point of expression for the -87kb enhancer, embryos were analysed at E12.5 and E13.5. At E12.5 the embryos expresses but not in the limbs (Figure 44A-C) but expression is confirmed in the somites. Sectioning reveals there are no chondrocytes or precursors expressing the transgene in the jaw (Figure 44G), ribs (Figure 44E) or limbs (Figure 44F) but staining is observed in the rear of brain only (Figure 44F), unfortunately sections deeper into the embryos were not obtained. A day later, at E13.5, definitive expression is seen in the whole mount, most noticeable is the rostral to ventral staining in the skin (Figure 45A). The developing joints do not stain at this stage but minor expression is seen in the limbs more proximal and towards the humerus and hip (Figure 45B and C). Histologically, there are no chondrocytes

expressing the enhancer, but in the fore paw there are a subgroup of cells that are active (Figure 45D) and this is seen more weakly in the hind limb (Figure 45G). The somites are expressing but not the ribs (Figure 45E). Other regions such as the nasal area does not express (Figure 45H) whereas the brain (Figure 45I) and eye (Figure 45F) presents with X-gal activity.

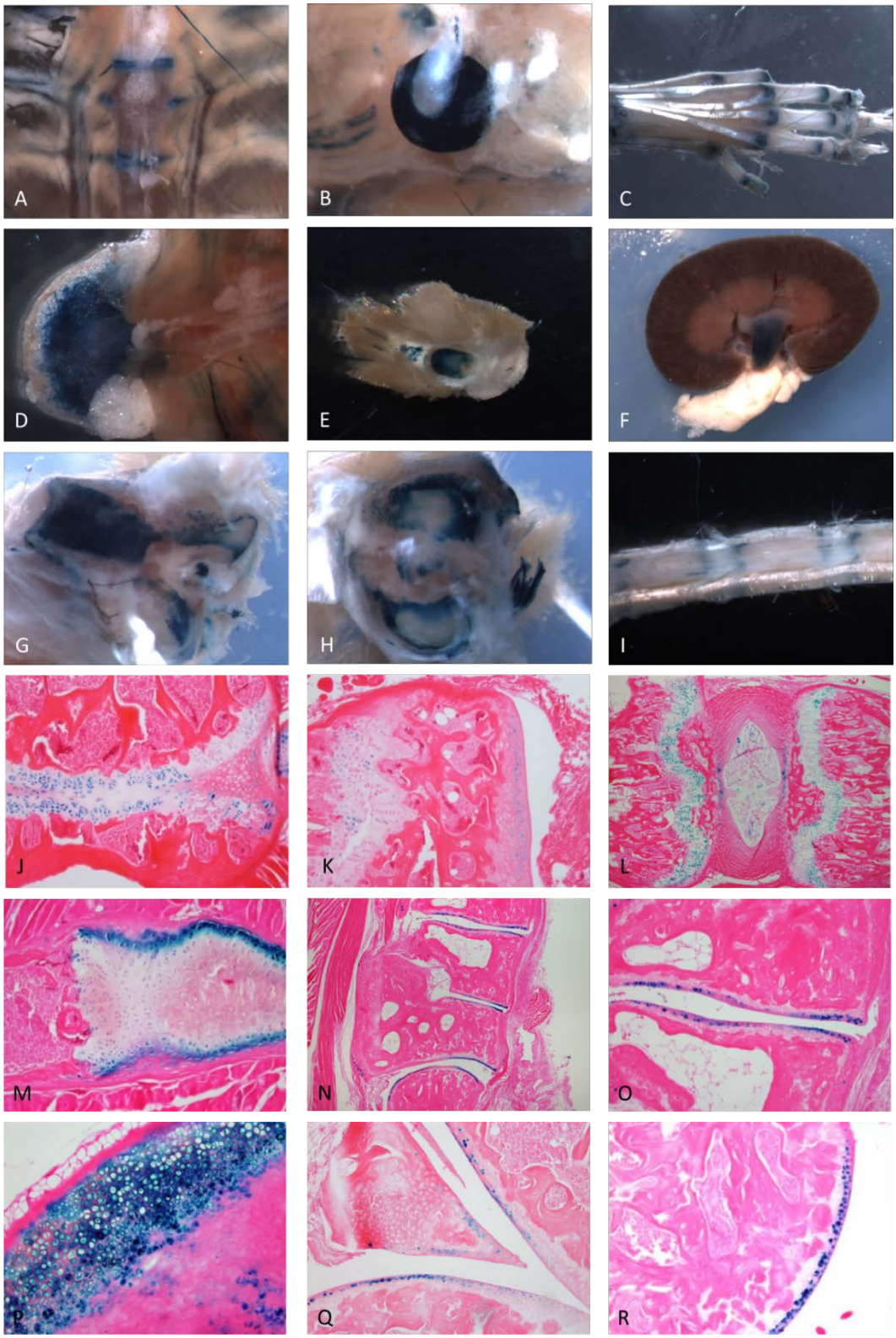
Figure 45: Whole mount and histological samples of the -87kb enhancer at E13.5



At E13.5 the enhancer (-87kb) starts to express in the limbs, as whole mounts shows that there is strong staining in the entire embryos with the exception of the external limbs such as the paws (A-C). In the limbs the expression is localised to a subset of cells not in

the chondrocytes that are forming the skeletal limbs and joints (D and G). The brain (I), somite's (E) and the eye (F) but the vibrissae do not (H).

Figure 46: 8 week old tissue samples and histology of the -87kb enhancer



8 week old tissue taken from the -87kb enhancer line and stained for B-gal activity. Whole mount of the ribs (A), femoral head (B), paw (C), Condyle (G), tibial plateau (H) and tail (I) shows specific staining in the joints and areas known to contain chondrocytes. Staining is also seen in the xiphoid (D) and the patella (E) but not in the kidneys (F). Histologically, the staining is seen in the superficial regions of the articular cartilage of the femoral head (P), knee and meniscus (O and K) (it must be noted that due to penetration and user error staining in these regions do not appear consistent and strong but these were the images chosen due to clarity) and tarsals (L and M). These is staining in the xiphoid (N), superficial region of the ribs (K), growth plates (J), the cartilage end plates of the IVD (J) but not in the nucleus pulpous (J).

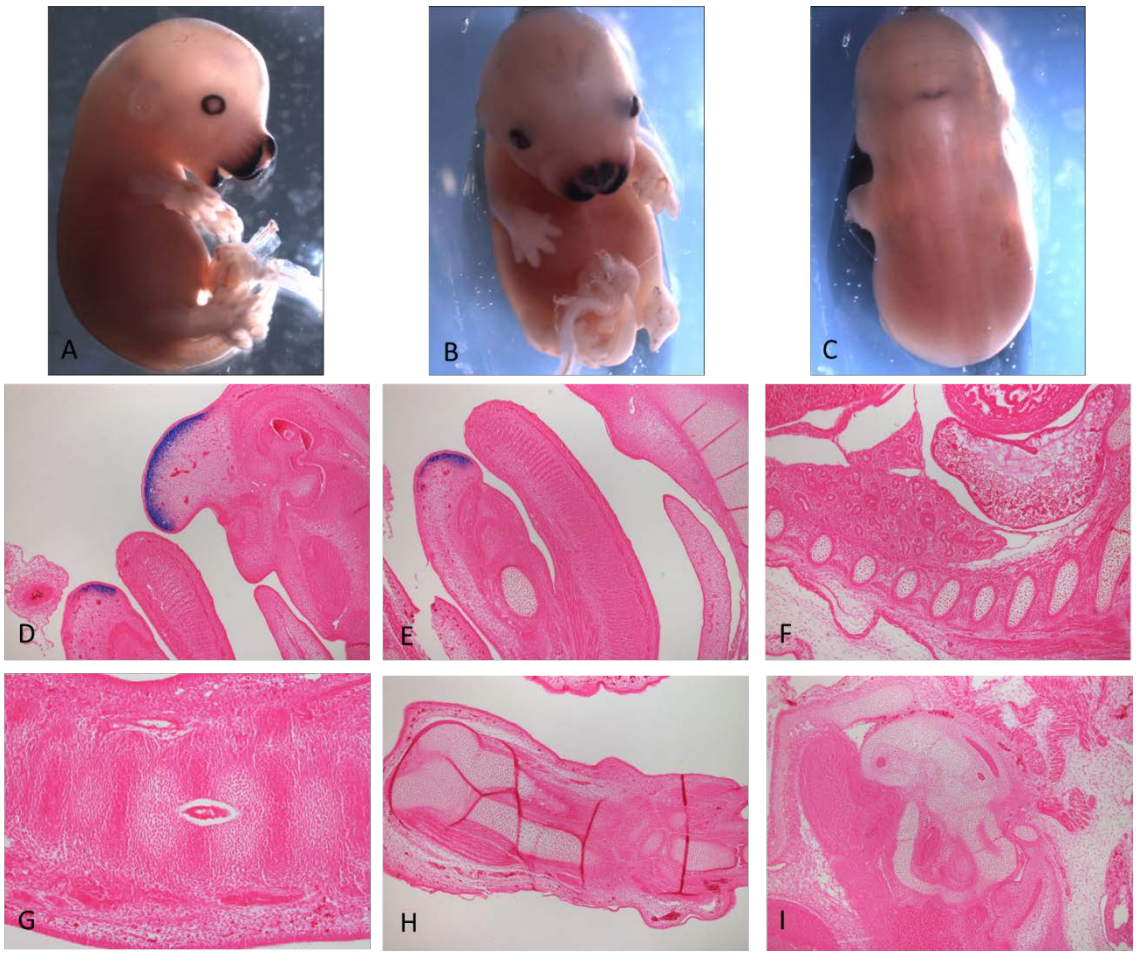
4.2.4. Exploration of the minimal sequence able to drive expression of the -87kb enhancer

The original sequence of the -87kb is large for an enhancer, 2.6kb, naturally we wanted to establish the minimal sequence capable of driving expression. Using conservation we cloned the opossum sequence to determine evolutionarily, how active the -87kb was, and this fragment was 1100bp. A shorter sequence of 754bp was generated from the most evolutionary conserved region. The fragments were cloned into the HSP68 vector and transient expression of transgenic mice were analysed at E15.5. The opossum sequence failed to express in chondrocyte specific areas but appeared on the snout of the mouse (Figure 47A-C). The expression when analysed histologically, showed no expression in chondrocytes in the limbs (Figure 47H), IVD (Figure 47G), ribs (Figure 47F), ear (Figure 47I) or in the Meckel's cartilage (Figure 47E) but expression was seen in the superficial region of the nose (Figure 47D).

The smaller fragment was able to express and was more defined in the ribs and limbs (Figure 48A and B) but due to skin staining it was difficult to establish the skeletal element staining. Histologically, the staining was chondrocyte specific, although more cells were stained than with the larger fragment it was still scattered and looked to

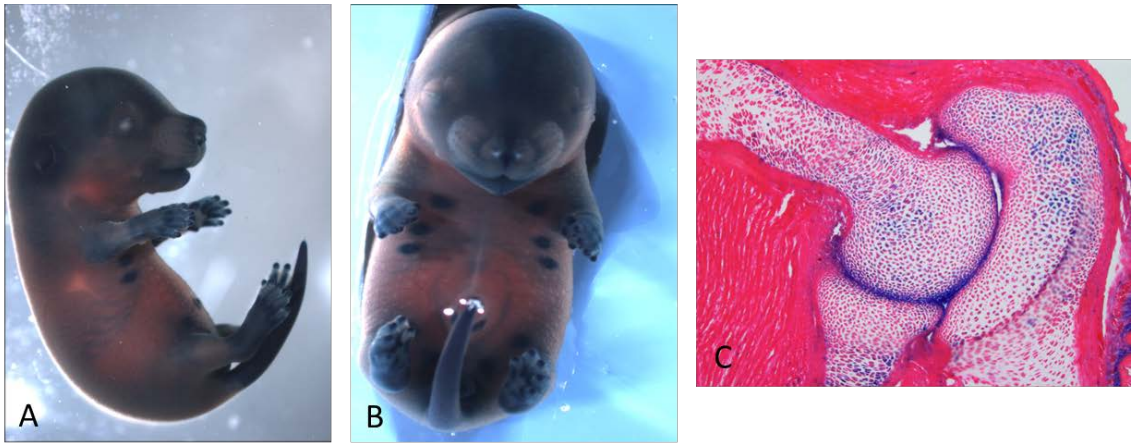
favour the joints (Figure 48C).Unfortunately, due to technical errors there was a lack of quality histological images to further identify the expression pattern of the smaller fragment, therefore it remains to be further investigated.

Figure 47: Opossum sequence of the -87kb enhancer is not able to express



Analysis of the opossum sequence of the -87kb enhancer. Transgenic mice only presented in whole mount with staining on the nose with minor staining in the back of the head (A-C). Histologically, there are no chondrocyte staining in the ribs (F), tail (G), forelimb (H), ear (I) or in Meckel's cartilage (E). The staining is apparently only in the superficial region of the mouth (D and E).

Figure 48: 700bp fragment of the -87kb enhancer is able to express



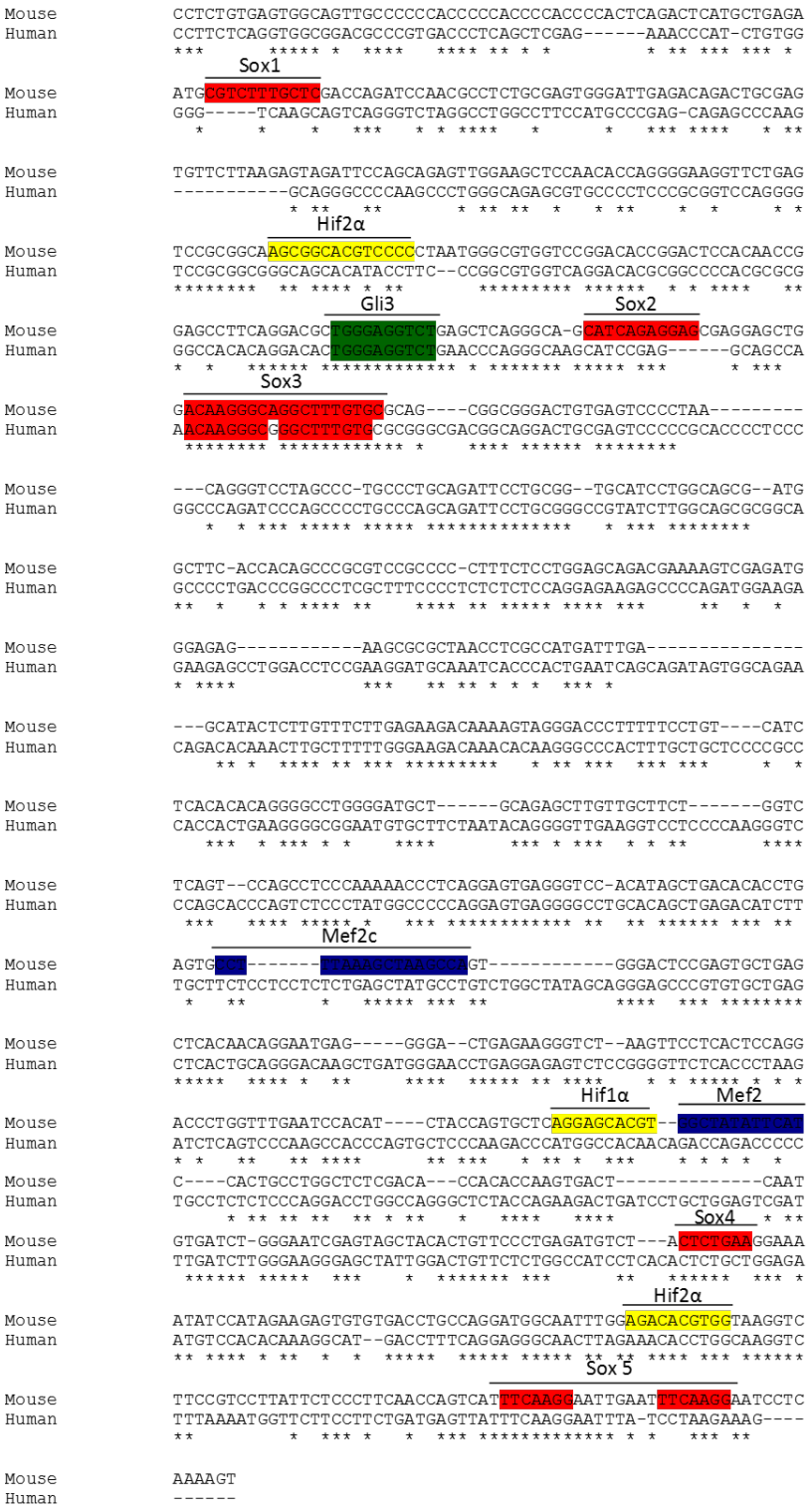
A 700bp of a smaller region of the -87kb was tested in transgenic mice. Transient transgenic embryos at E15.5 showed strong staining in skeletal elements, (A) shows a representative embryos. Due to sectioning issues there were poor attainment of presentable histological images, (C) shows the humerus, radius and ulna with similar scattered pattern of expression of the whole sequence of the -87kb.

4.2.5. Identification of transcription factor binding regions of the -65kb and -87kb

To establish which TFs interact with the -65kb and -87kb enhancer to bring about their unique express, each sequence was aligned to other species using ClustalW to help identify conserved TF binding sites. The -65kb was not as well conserved, as the sequence was only able to align to the human sequence. The -87kb sequence was better conserved as it aligned up to opossum seen in Figure 49 and Figure 50. Using Biobase multiple Sox9 binding sites were found in the -65kb and -87kb elements, there were 5 sites in the -65kb and in the -87kb there was 10 sites, in both there only one site containing the paired adjacent Sox9 binding site (Bridgewater et al., 2003). There were other conserved TF binding sites that were of interest, such as the Mef2, Hif and Gli proteins in the -65kb which may regulate the hypertrophic chondrocyte specific expression. In the -87kb there were notch effectors such as rbpj-k, age related proteins

such as Nfat and other such as SP1, SP3, Hifs, AP1 and TWIST1. For EMSA experiments only Sox9, Nfat and rbpj-k was examined.

Figure 49: ClustalW alignment of the -65kb enhancer and possible transcription factor binding sites

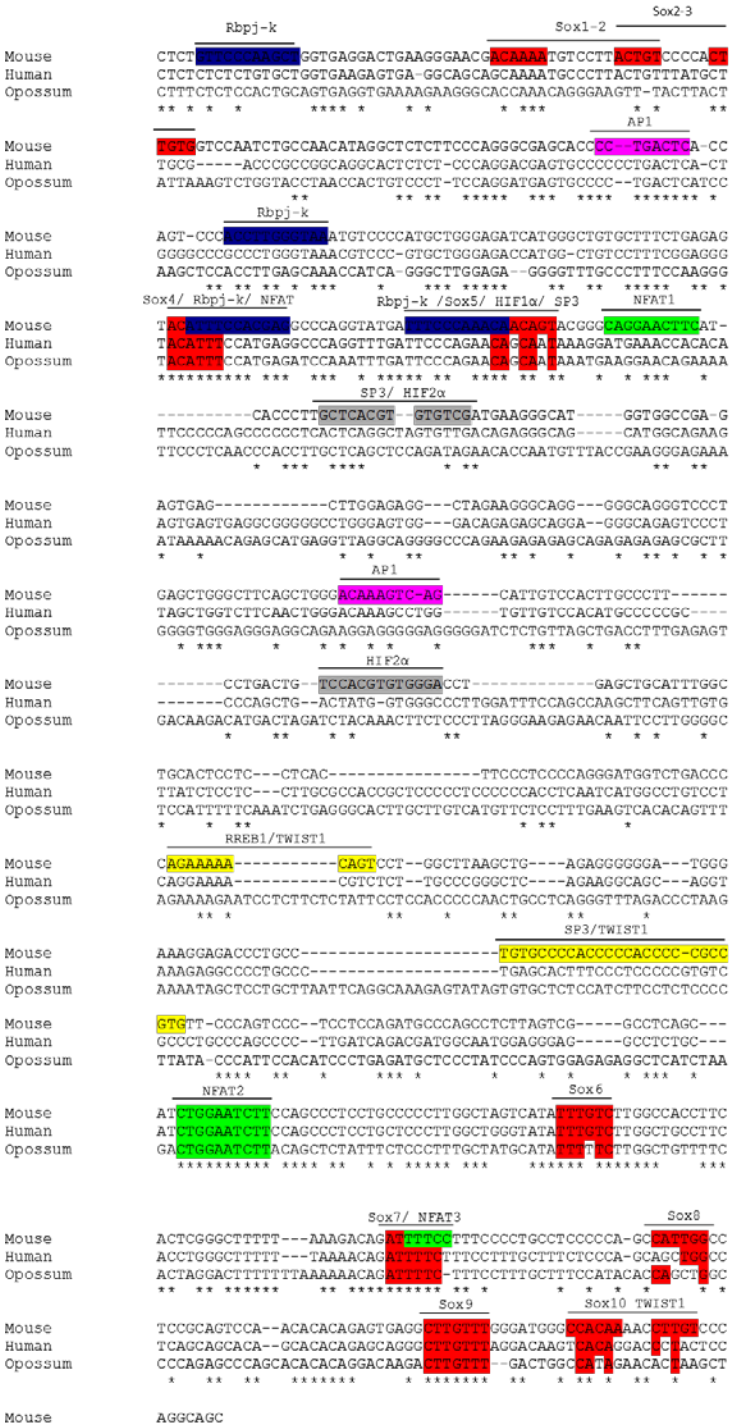


ClustalW alignment of the -65kb enhancer between human and mouse. Transcription factors identified using Biobase transfec was annotated. Sox9 binding sites are shown in red numbered 1-5. Hypoxia -inducible factor (Hif) are shown in yellow. Myocyte enhancer factor 2 (Mef2) are shown in blue. The zinc finger protein Gl3 is shown in green. Only Sox9 binding sites were used in EMSA experiments.

Indirect competitive EMSA experiments with Sox9 protein and Sox9 control probe revealed that in the -65kb one area (site 3) interacted *in vitro* with Sox9 and mutations introduced into the binding sequence resulted in a loss of binding (Figure 51A). Site 5 also produced a weaker signal compared to control, although a band is still seen but mutations failed to diminish this binding. The mutated sequence did not produce a Sox9 binding site and there were no remnants of a binding site. In the -87kb there are 3 binding sites revealed by EMSA, sites 1-2, 5 and 9 and mutations in these sites removed the binding (Figure 51B).

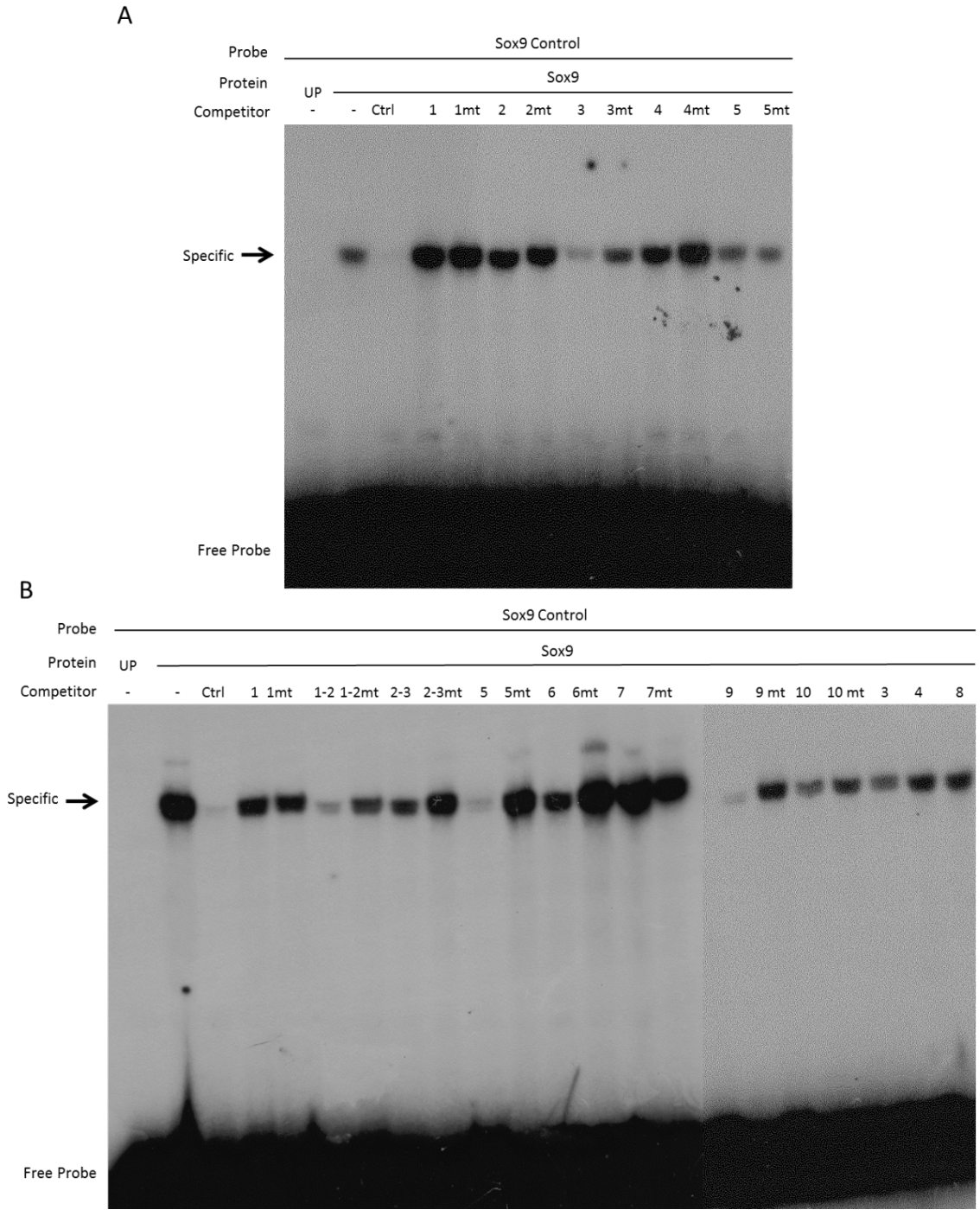
For the -87kb enhance Nfat binding was examined by EMSA due to its suggested role in adult cartilage and age related regulation (Rodova et al., 2011). Of the 3 predicted sites in indirect competitive assays all the sites bound and mutations abrogated the affinity of the sites (Figure 52). Although mutated fragments were generated by GeneArt™ string for the -87kb Sox9 and Nfat binding sites and cloned into the expression vector for *in vivo* analysis, due to time and financial constraints transgenic mice were not generated. Therefore, the importance of these TFs and their interactions with the -87kb remains to be explored.

Figure 50: ClustalW alignment of the -87kb enhancer and possible transcription factor binding sites



ClustalW alignment of the conserved regions of the -87kb, the mouse sequence is aligned to human and opossum. Sox9 binding sites are shown in red, Rbpj-k is shown in blue, green marks the Nfat binding sites, TWIST1 is highlighted in yellow, Hif in grey and Activator protein 1 (AP1) in pink. Rbpj-k, Sox9 and Nfat binding sites were used for EMSA experiments.

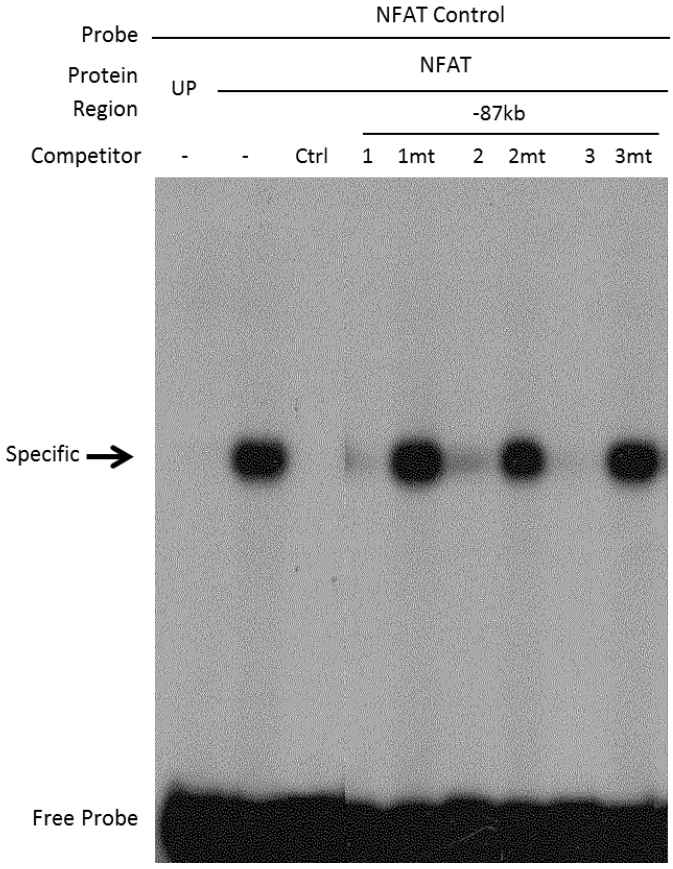
Figure 51: Indirect competitive EMSA for Sox9 in the -65kb and -87kb enhancer



Indirect competitive Electromobility shift assay using the Sox9 to detect interactions within the -65kb and -87kb enhancer. The Sox9 control oligo was used as the probe and an unlabelled oligo of the same sequence is the control competitor. Lane 1 UP is the

lysate from the in vitro translation reaction with no vector to determine any non-specific binding. The second lane is the probe with no competitor to show it is binding to the proteins. The third lane is the control competitor therefore the band should be removed if it is able to bind. Sox9 interacts in the -65kb in vitro only at one site (site 3) where as in the -87kb there are 3 binding sites, (1-2, 5 and 9). Mutations in these sites removes the binding seen by the mutation copy (mt).

Figure 52: Indirect competitive shift assay for Nfat binding in the -87kb enhancer



Indirect competitive Electromobility shift assay using the NFAT core protein binding to detect interactions with the -87kb enhancer. The NFAT control oligo was used as the probe and an unlabelled oligo of the same sequence is the control competitor. Lane 1 UP is the lysate from the in vitro translation reaction with no vector to determine any non-

specific binding. The second lane is the probe with no competitor to show it is binding to the proteins. The third lane is the control competitor therefore the band should be removed if it is able to bind. In vitro Nfat interacted with all 3 identified binding sites and mutations (mt) in those sites removed the binding.

4.3. Analysis of the 3' intronic enhancers of Acan

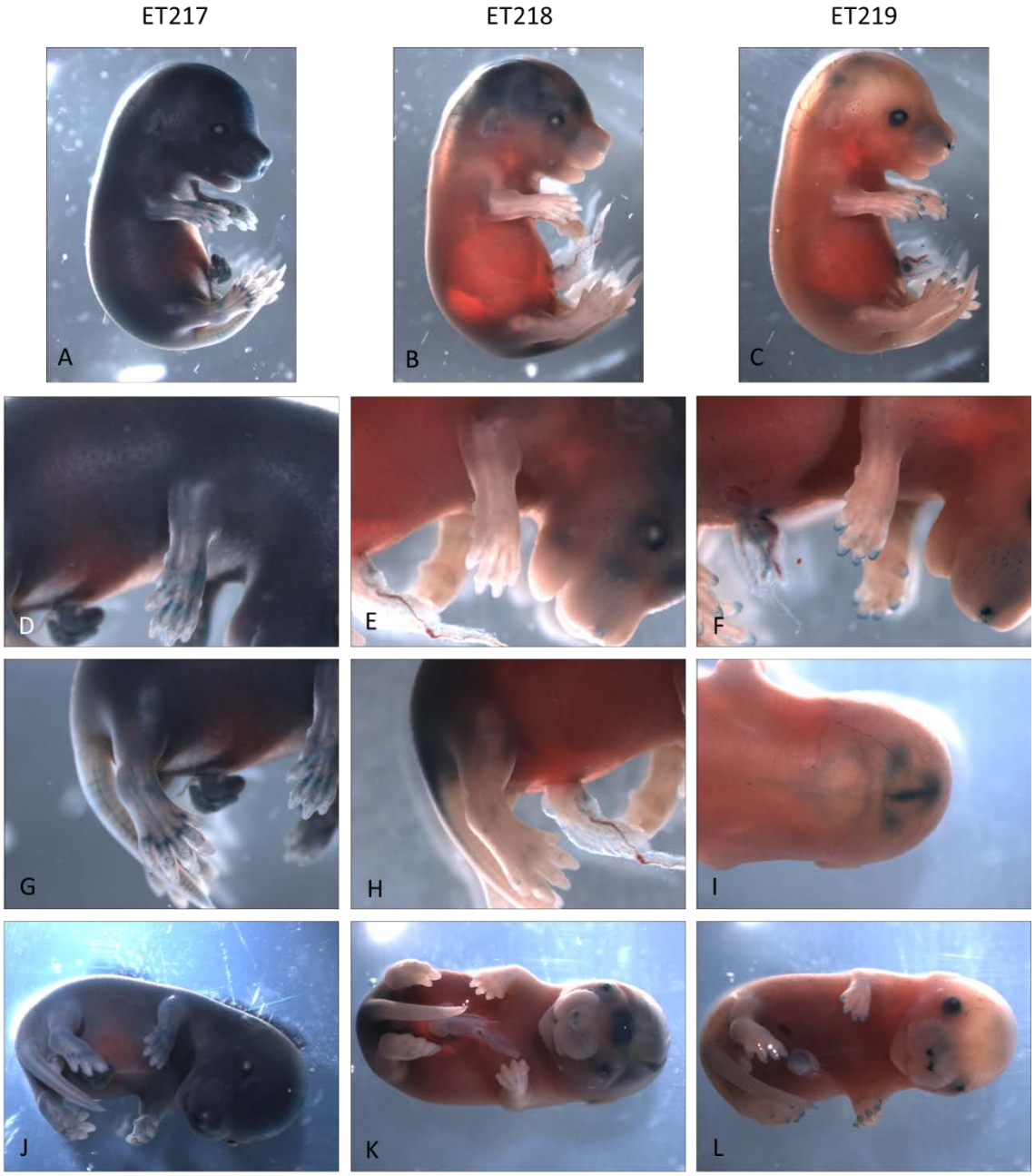
4.3.1. Background

Enhancers do not only reside within the intergenic regions, regulatory elements closer to the 3' end of a gene are able to drive strong expression of target genes or drive splicing of gene isoforms (known as intronic splice enhancers) (Wang et al., 2012). Recent loss of DNA methylation studies have shown that in human colon carcinoma (HCT116) and DKO1 (HCT116 cells with DNMT3b knocked out) showed that the majority of enhancers reside in the introns of genes (Blattler et al., 2014). The importance of these intronic enhancers were interrogated in *Gli3* where 6 enhancers located in the second, third, tenth and thirteenth introns drives limb specific and central nervous system expression (Anwar et al., 2015, Abbasi et al., 2010). Additionally, in the vascular endothelial growth factor receptor Flk1 an enhancer in the tenth intron is able to drive expression of reporter constructs in arterial endothelial cells, starting as a pan-vascular expresser and then becoming arterial restricted in mouse and zebrafish (Becker et al., 2016). This enhancer is under the influence of Gata and Ets TFs and Rbpj-k restricts the expression to arterial cells. Therefore, we screened the 3' intronic sequences of *Acan* but there were few limb specific histone marks for enhancers in these regions but areas of high conservation. *Acan* is not only expressed in the cartilage, there is detectable expression in the heart valves and brain which were rarely seen in the intergenic enhancers identified, so we reasoned that the areas of conservation may house other tissue specific enhancers.

4.3.2. The last intron

Within the last intron of *Acan* the ENCODE analysis showed that it contained a small H3K4me1 but no other histone marks or DNase1 hypersensitivity to mark an enhancer. However, it was very well conserved in Fugu but not present in Zebrafish or chicken. Due to its conservation we believed that it may have a function, therefore it was cloned into the HSP68/LacZ vector and examined in E15.5 mice. When we examined the embryos the staining presented with 3 different patterns (Figure 53A-C). This suggests that, although the DNA fragment that was cloned is able to act as an enhancer due to its ability to drive the report, it may be influenced by the environment it was inserted into. The first expression (ET217) was pan- skin up until the distal limbs where staining is observed in the joint area (Figure 53A, D, G and J). There were gradual reduction in the skin expression with more craniofacial expression (Figure 53E and K) and weaker joint expression (Figure 53E and H) in ET218. The final expression, ET219, was mainly at the tips of the paws (Figure 53F), nasal (Figure 53F and L) and rear of the skull (Figure 53I). All three expression patterns do not appear to localise to skeletal elements, which suggests the enhancer may not be chondrocyte specific or active in the developing skeleton. To validate this theory histological examination was carried out.

Figure 53: Whole images of the variable expression pattern of the +60kb region

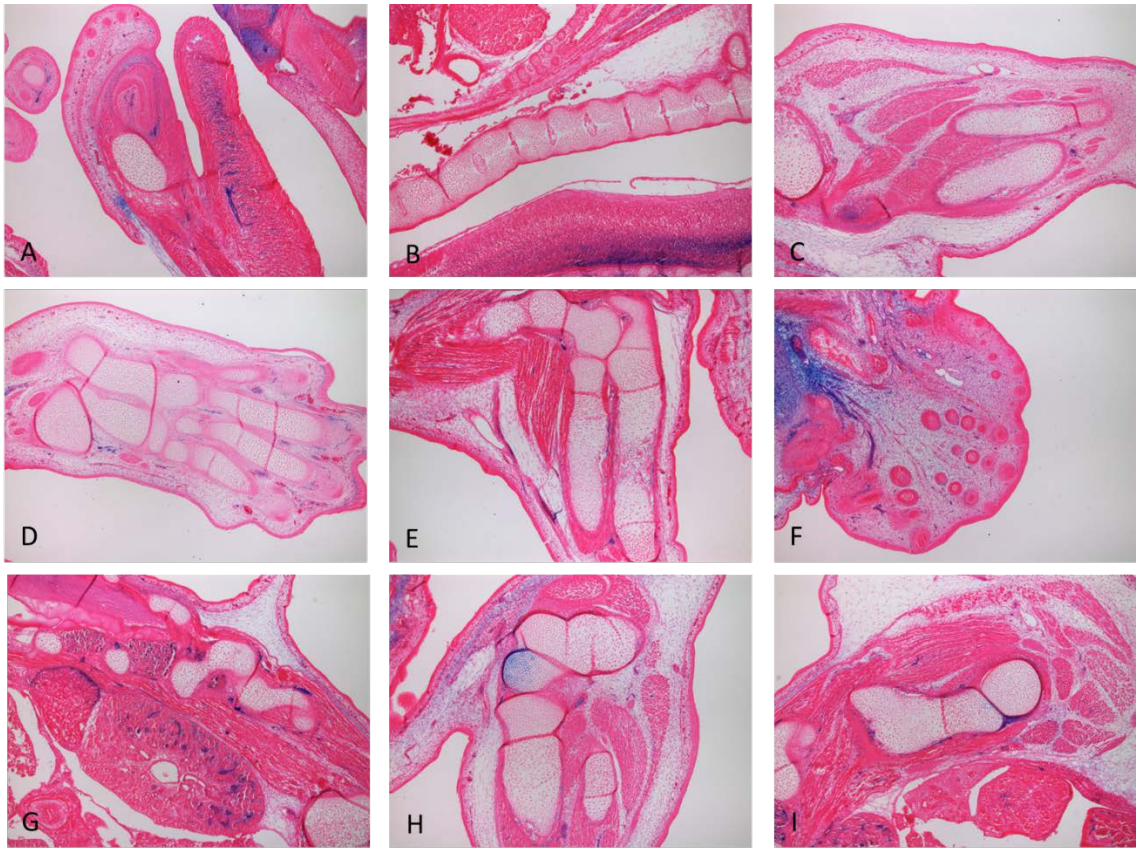


Whole mount images of the different expression patters of the +60kb regulatory element. ET217 showed pan- skin expression (A and J) with areas in the limbs (D) and tail (G) resembling the joint staining. ET218 showed less expression in the skin (B), the limbs (E) showed weak patterning particularly in the hindlimb (H), the craniofacial showed LacZ activity. ET219 showed minimal skin staining and no skeletal element activity (C).

On the limbs at the tips of the distal phalanges (F) and the nasal showed blue stain deposition (L). The rear of the skull stained along the cranial sutures.

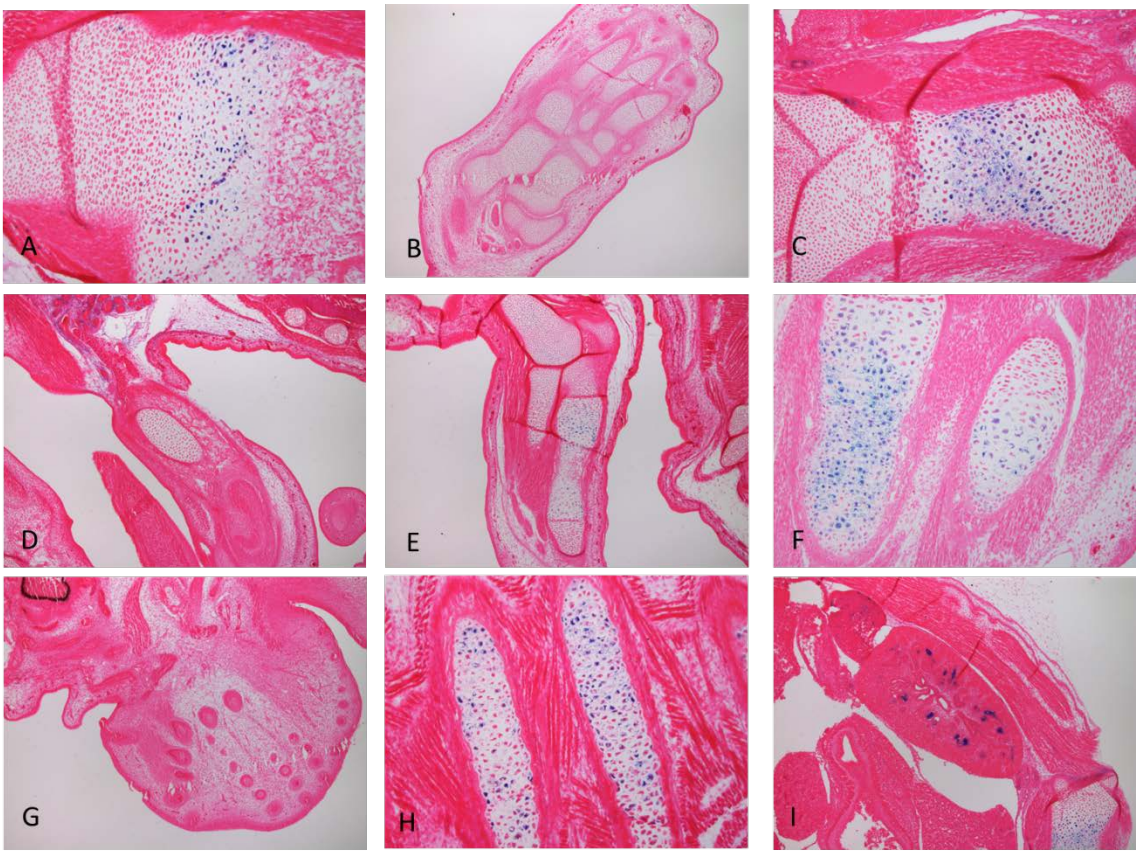
Histologically, the three patterns were widely different from each other sharing very few similarities, hinting that the region we cloned was too large, too short, missing an insulator region or other core regulatory elements. If the region was too large there is a chance that multiple elements were present that can produce multiple effects, if the sequence was too small, key sequences for TFs may be missing resulting in different proteins regulating the same sequence. One other factor is the position effects, the sequence may have insulator regions that causes it to disrupt the local environment that the transgene inserted into (Franke et al., 2016). There was only one expression pattern that expressed in chondrocytes, in ET218, there were weak expression in the core chondrocytes of the long bones such as the clavicle (Figure 55A and C), ribs (Figure 55H) radius and ulna (Figure 55F) but the pattern is weak and scattered (Figure 55E). There was noted expression in the kidneys similar to the +26kb (Figure 55I) but no expression in the paws (Figure 55B), Meckel's cartilage (Figure 55D) or vibrissae (Figure 55G). The pan-skin expression staining, ET217, resided in the fibroblasts and surrounded the forming bones of the paw (Figure 54D), radius, ulna (Figure 54C) and femur (Figure 54I), but no chondrocytes appear to be staining in the limbs (Figure 54E), IVD (Figure 54B) or the femur itself. However, there was a patch of staining in the chondrocytes seen in the femur when sectioned adjacently (Figure 54H). There were no staining in the vibrissae (Figure 54F), Meckel's cartilage (Figure 54A) and again staining was seen in the kidney (Figure 54G). The weaker expression in ET219 showed spinal cord staining (Figure 56A), nasal (Figure 56C), regions in the ear (Figure 56H) and in the nucleus pulposus of the tail (Figure 56E) but no chondrocyte expression in the femur (Figure 56I), bones of the ear (Figure 56H), paw (Figure 56D), vertebra (Figure 56A), fore limb (Figure 56G) or in Meckel's cartilage (Figure 56B). This embryo did not present with staining in the kidney (Figure 56F) unlike the other.

Figure 54: Histology sample of ET217 of the +60kb enhancer



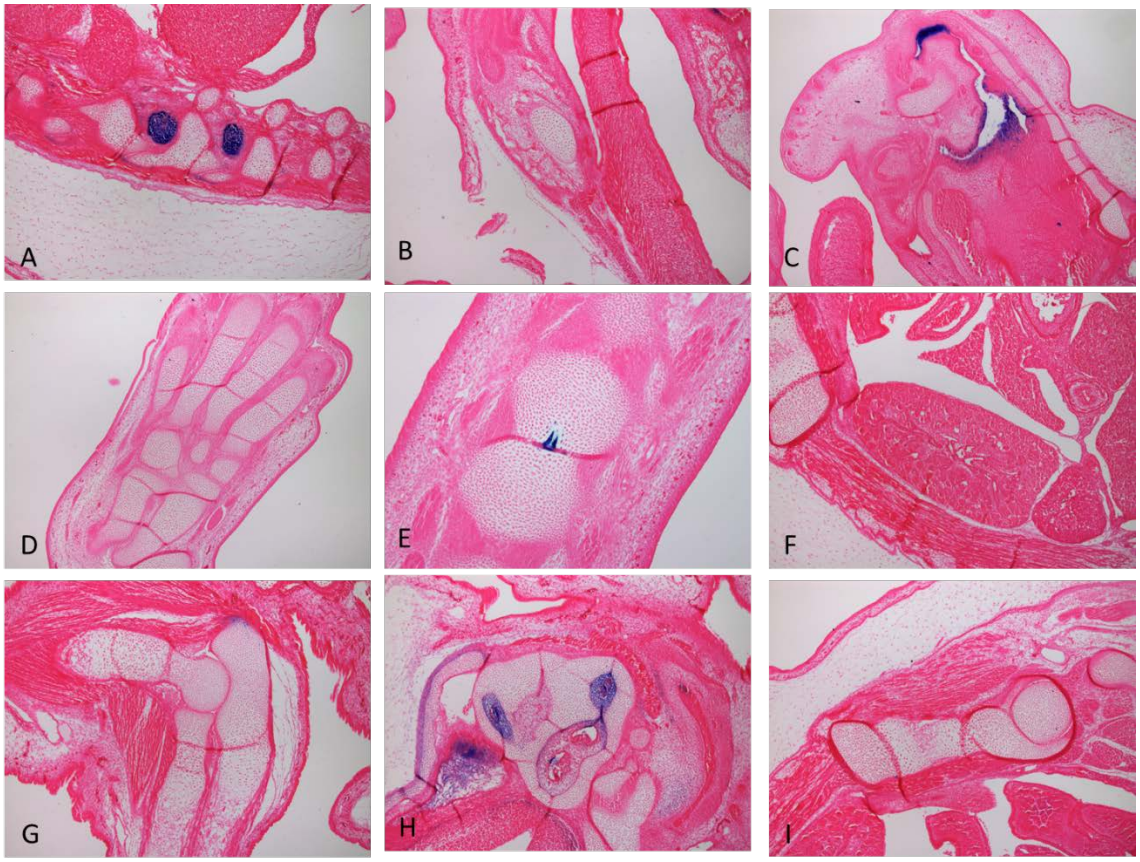
Sections of the ET217 embryos with no chondrocyte specific staining in the metacarpals, phalanges (D), radius, ulna or humerus (C and E). In the femur the staining surrounds the forming bone (I) but in a subset of chondrocyte of the femur there was localised expression (H). The staining was seen in the fibroblasts of the nose (F) and jaw (A), and around bones (D and C) but not in the Meckel's cartilage (A). There was also noticeable staining in the kidney residing in the Bowman's capsule (G) but no staining in the vertebra.

Figure 55: Histological examination of the ET218 embryo of the +60kb enhancer



Sections of the ET218 of the +60kb element embryos showed chondrocyte staining in the pre- hypertrophic cells in the clavicle (A and C) and radius and ulna (F and E) and in a scattered pattern in the ribs (H), but low magnification of the fore limb shows very weak expression of the LacZ (E), there was no staining in the paw (B), Meckel's cartilage (D) or vibrissae (G).There was noticeable staining in the kidney residing in the Bowman's capsule (G).

Figure 56: Histological examination of ET219 of the +60kb enhancer



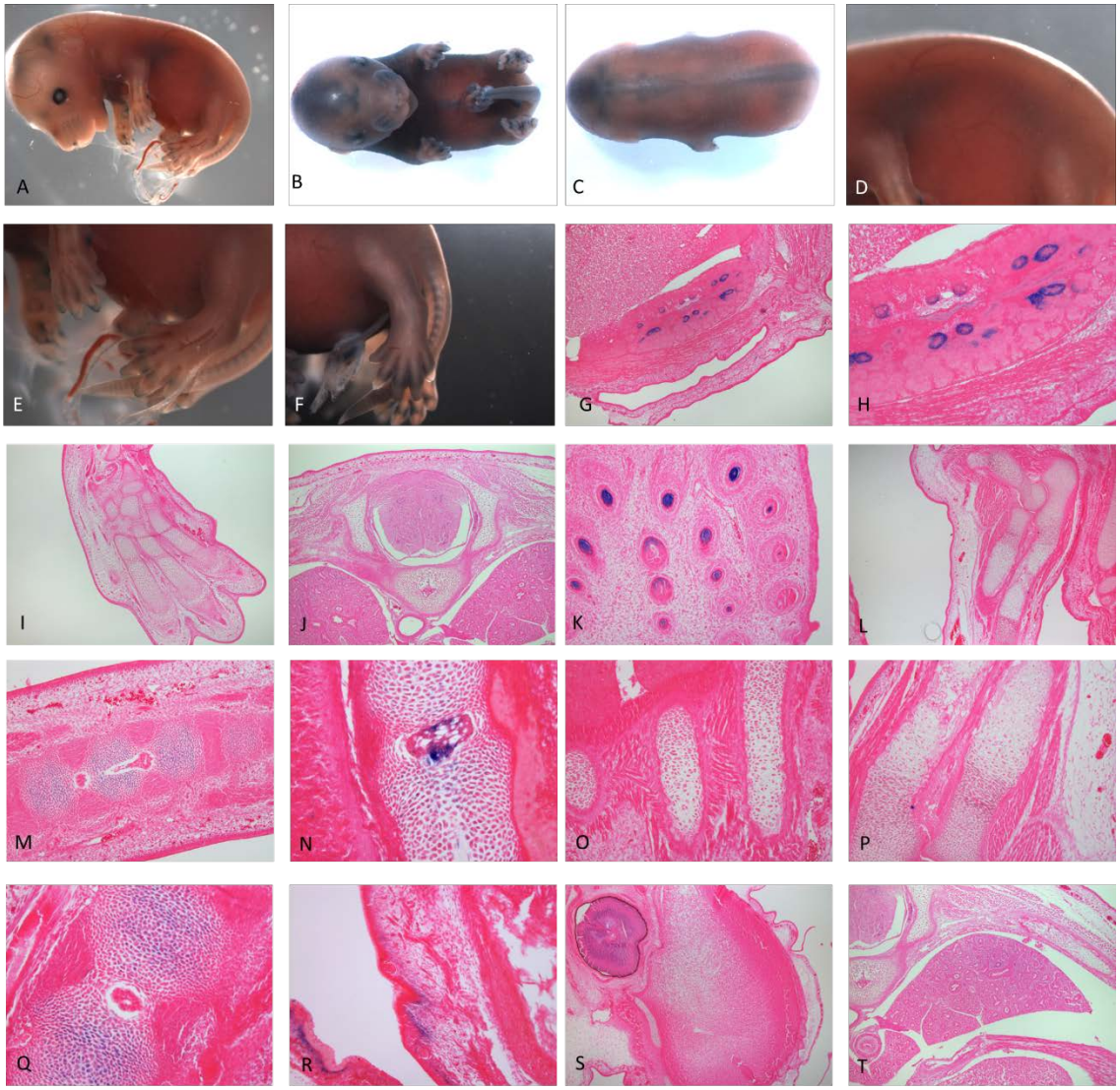
Sections of the ET219 embryos of the +60kb with no chondrocyte specific staining in the metacarpals, phalanges (D), the vertebra (A), Meckel's cartilage (B), femur (I), skull (C) radius, ulna or humerus (D).There is no staining in the kidney or the chondrocytes in the tail (E) but staining is seen centrally in the nucleus pulposus. Staining in the ear is not localised to the chondrocytes but rather the membrane (H) similar to the nasal cavity (C). The spinal cord is seen to express the transgene (A).

4.3.3. Enhancer in the 12th intron of Acan

Due to evolutionary conservation a 2.5kb region of the 12th intron that had minor histone marks for H3K4me1 in the limb indicating a poised enhancer was cloned into the HSP68/LacZ vector to be examined at E15.5, this was named +55kb. Whole mount revealed that there was enhancer activity although it did not look cartilage specific

(Figure 57A-C). Staining seemed to line the edges of joints (Figure 57E), the spinal cord (Figure 57C) but not the vertebra (Figure 57D) and areas of the skull (Figure 57B). There was staining the tail seen clearly (Figure 57E and F) that resembles the pattern of the annulus fibrosis. Histologically, there were no chondrocyte staining seen in the paw (Figure 57G) and vertebra (Figure 57H). In the tailit appeared to mark the chondrocytes (Figure 57J) but in the IVD of the lumbar spine activity of the enhancer was on the edge of the nucleus pulposus (Figure 57K), there was noted staining in the peritubular interstitial cells of the kidney (Figure 57I and L).

Figure 57: Whole mount and histological images of the +55kb enhancer

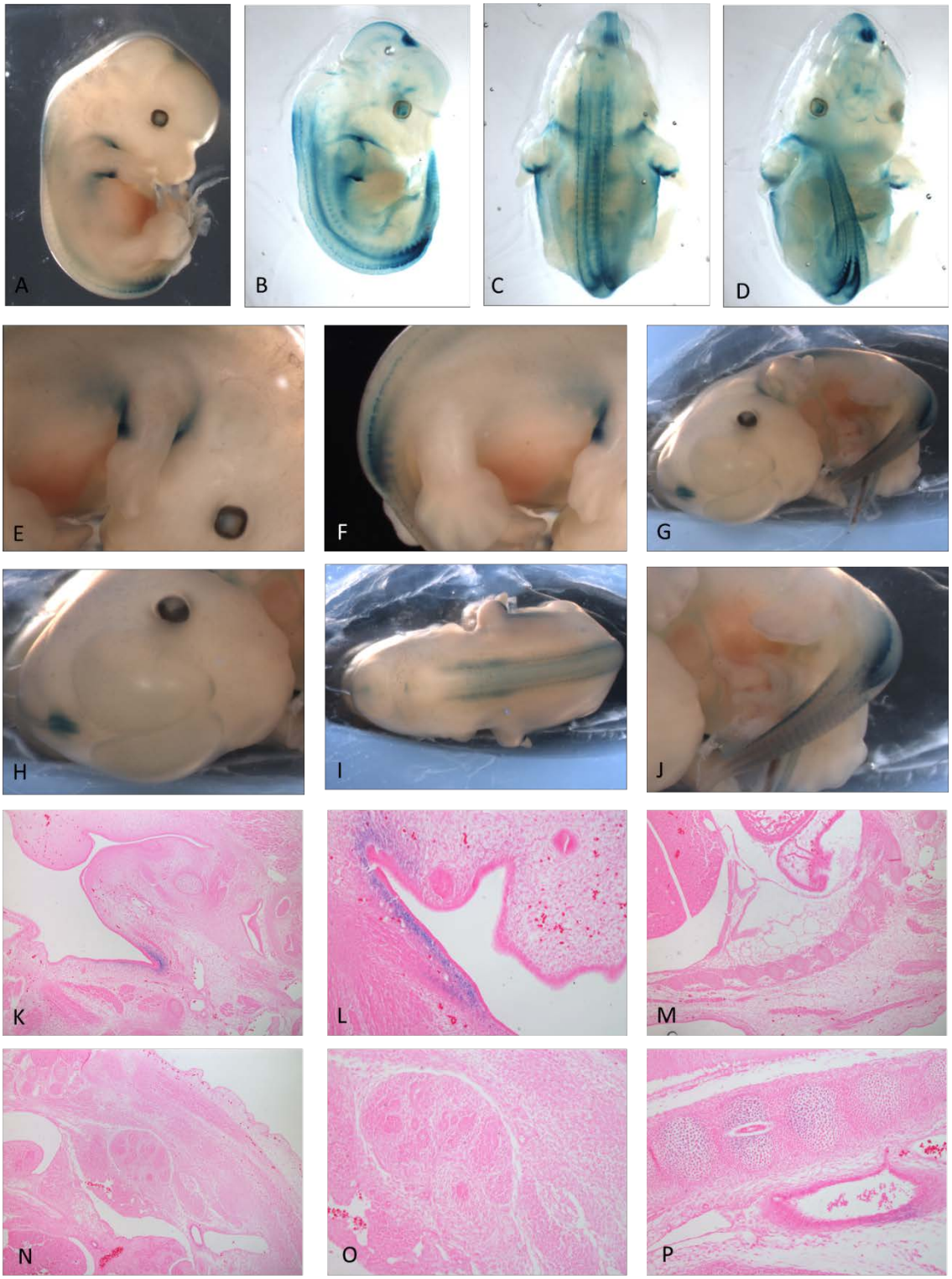


Whole mount images of the +55kb regulatory activity at E15.5 shows staining in the limbs, tail, skull and in the spinal column (A-C). Higher magnification shows no observable vertebra staining (D) but in the limbs staining is localised to the joints and the central regions of the tail (E and F). Histologically, there is no chondrocyte staining in the paw (I), spine (J) but some in the tail (M and Q). Staining is seen on the edge of the nucleus pulposus of the lumbar spine (N) and in the kidney (G and H). Transverse sections show weak staining in the lungs (T) but no staining in the limb (L and P) or ribs (O).There

is staining in the vibrissae (K) and in the epidermis of the skin (R) and possibly in the eye but not in the brain (S).

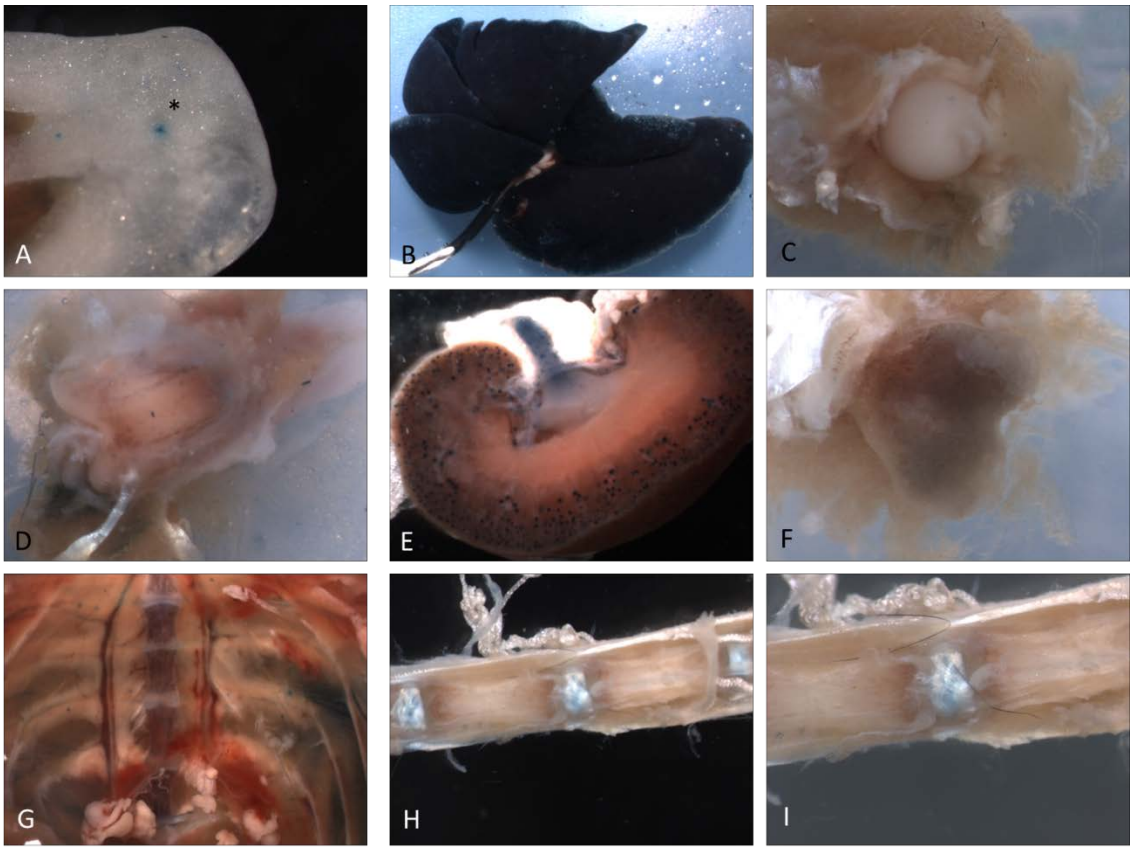
To determine if the expression changed at earlier time point in development or postnatally, embryos were collected at E12.5 and tissue was collected from 8 week old mice. Whole mount images of E12.5 embryos showed staining in the spinal cord (Figure 58E and F), a central area of the skull (Figure 58D) and regions under the forelimb (Figure 58A and B) but not in the hindlimb (Figure 58C) following developmental process of expanding rostral –caudally and dorsal – ventrally. In 8 week old tissue cartilage tissue did not stain for X-gal such as the condyle (Figure 59D) and plateau (Figure 59F) of the knee, hip (Figure 59C) and ribs (Figure 59G). However, in tissues, such as the lungs (Figure 59B), there was strong staining and in the kidneys it was reserved for the Bowman's capsule (Figure 59E). Additionally, in the xiphoid there was some staining observed (Figure 59A) a cartilaginous tissue which leaves the question of what connects these different tissue types that allows this enhancer to function. . In the tail staining did not localise to the cartilage end plates but rather in the central zones where the annulus fibrosis and nucleus pulposus reside (Figure 59H and I). Upon examining histologically the chondrocytes did not stain in the ribs (Figure 60A), growth plate of the sternum (Figure 60D), articular cartilage of the femoral head (Figure 60D), or knee (Figure 60E and F). On some occasions ectopic staining was seen in single cells in the ribs (Figure 60C). When examining the IVD and the tail we see that there is no staining in the cartilage end plate, growth plate, or annulus fibrosis, but centralised in the nucleus pulposus (Figure 60G, H and I). Other non-cartilage based tissue staining was localised in the lungs (Figure 60K) and in the inner lumen of blood vessels (Figure 60L) as well as in the peritubular interstitial cells in the kidneys (Figure 60J). The staining in the kidneys remain even at 6 months of age whereas the lungs intensity of staining was reduced but still present, but no staining was in the articular cartilage (Figure 61).

Figure 58: E12.5 whole mount and histological examination of the +55kb enhancer



Whole mount images of the +55kb element at E12.5 presents with no limb staining (A) but staining in the spine (G, F, I and J) and localised staining at the proximal region of the fore limb (E and F) but not in the hind limb (F), there is also a localised spot on the head (H and G) that presents the stain deposition. Clearing the embryos with KOH show no skeletal element staining and more localised staining in the spinal cord (B-D). Histological examination show that the staining is in the epidermis (K and L) but no staining in the ribs (M), kidney (N and O) or IVD (P).

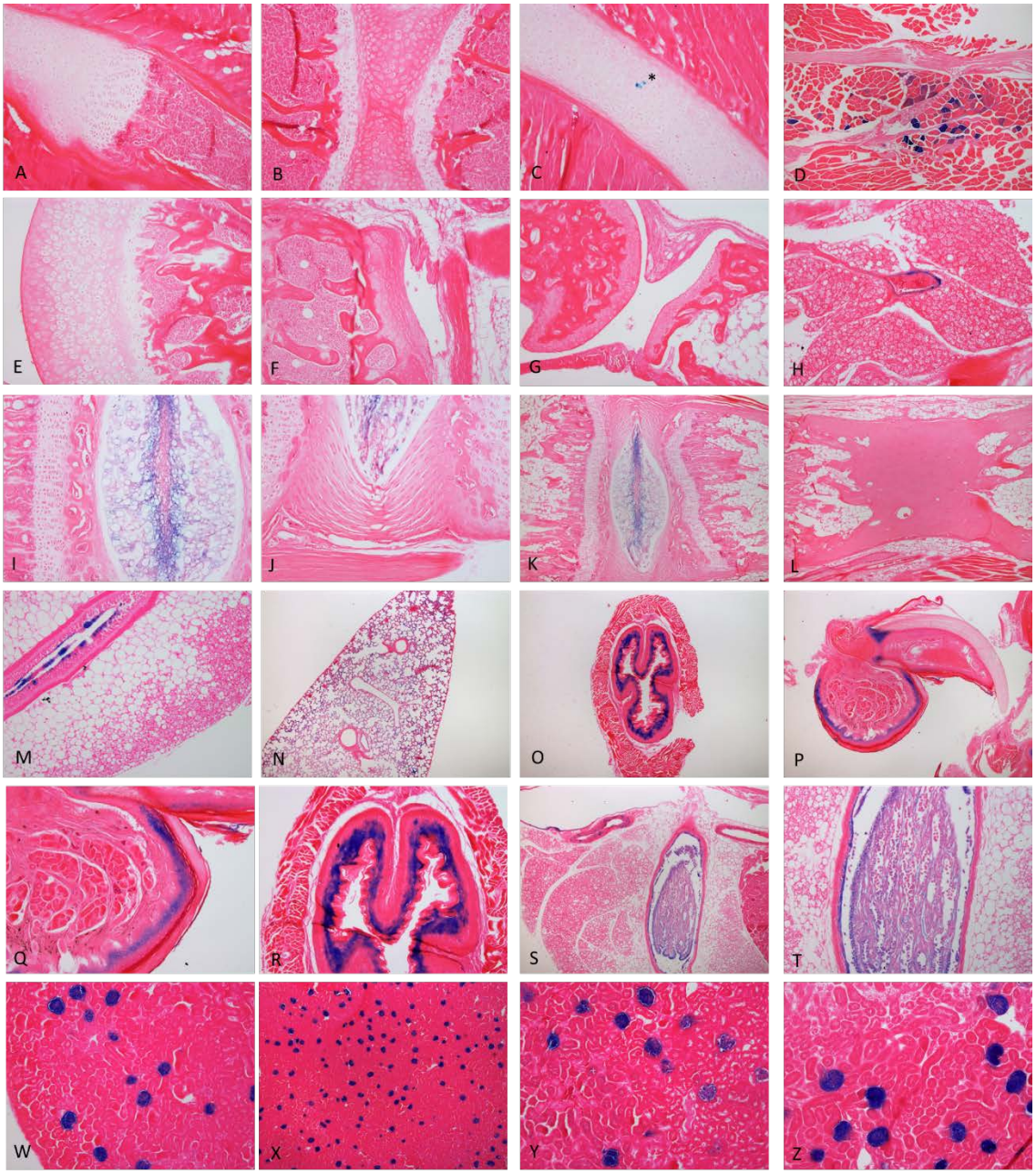
Figure 59: 8 week old tissue whole mount images of the +55kb enhancer



At 8 weeks of age the +55kb whole mount tissue images show none of the cartilage of the femoral head (C), condyle and plateau of the knee (D and F) or ribs (G) have any transgene activity. But in the lungs (B) and kidneys (E) there is B-gal activity. The xiphoid

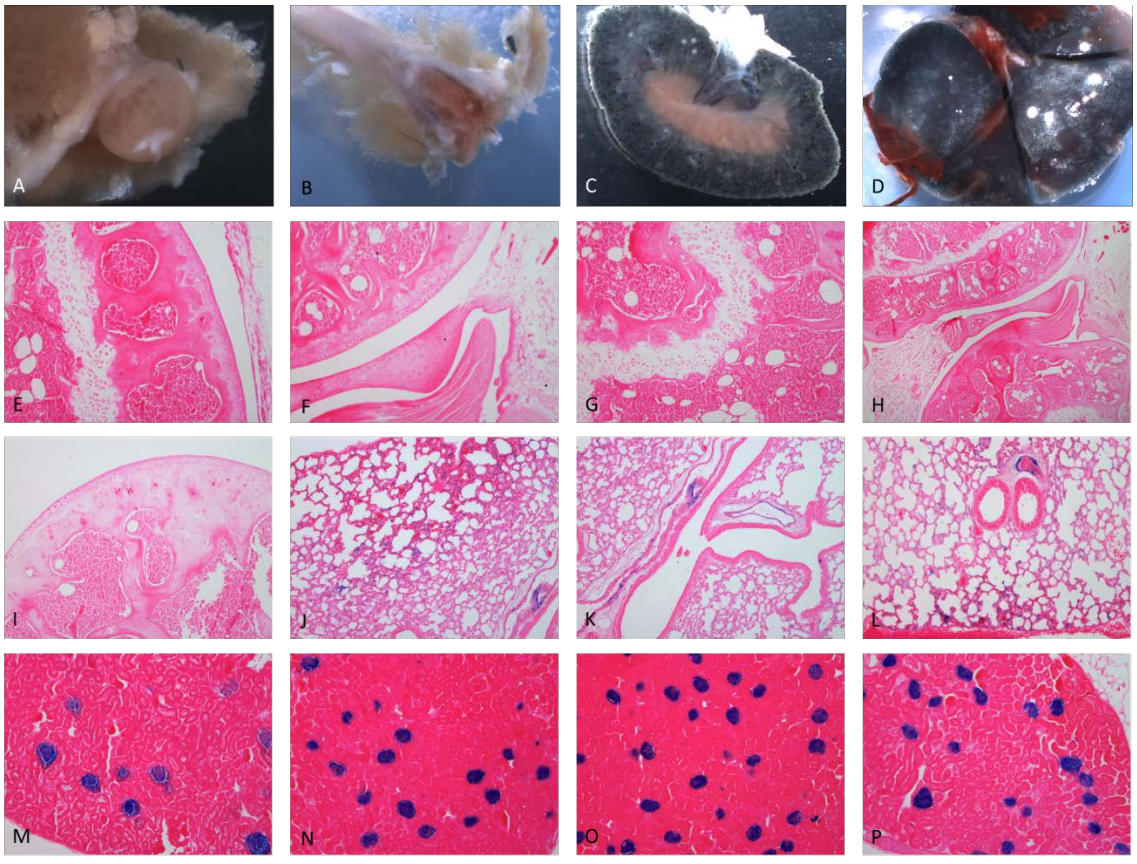
(A) has ectopic staining. In the rain there is blue staining seen in the central region of where the annulus fibrosis and nucleus pulposus reside (H and I)*

Figure 60: Histological examination of the -55kb enhancer in 8 week old tissue



Histological examination of tissue at 8 weeks of age show that there is no chondrocyte staining in the ribs but one cell [*] (A and C), growth plate of the sternum (B) or in articular cartilage of the femoral head (E) or knee (F and G).In the IVD there is no staining in the growth plates or annulus fibrosis but in the central area of the nucleus pulposus there is X-gal activity (I, J and K) but not in the bones (L). In tissue that do not contain cartilage there is staining in the peritubular interstitial cells of the kidney (W, X, Y and Z), in the lungs (N), the basal lumen of the skin (P and Q), renal (S and T), in the lumen of the aorta (O and R) and other blood vessels (H and M).

Figure 61: 6 month old tissue images of the +55kb enhancer



Whole mount of tissue at 6 months of age show that there is no chondrocyte staining in the hip (A), or knee (B) but blue is seen in the kidney (C) and lung (D). Histological examination of the knee show there is no staining in the growth plates or articular

cartilage (E-H) or in the femoral head (I). In tissue that do not contain cartilage there is staining in the peritubular interstitial cells of the kidney (M-P) and in the lungs (J-K).

4.3.4. Antibody staining to determine Acan expression and HAPLN3 expression

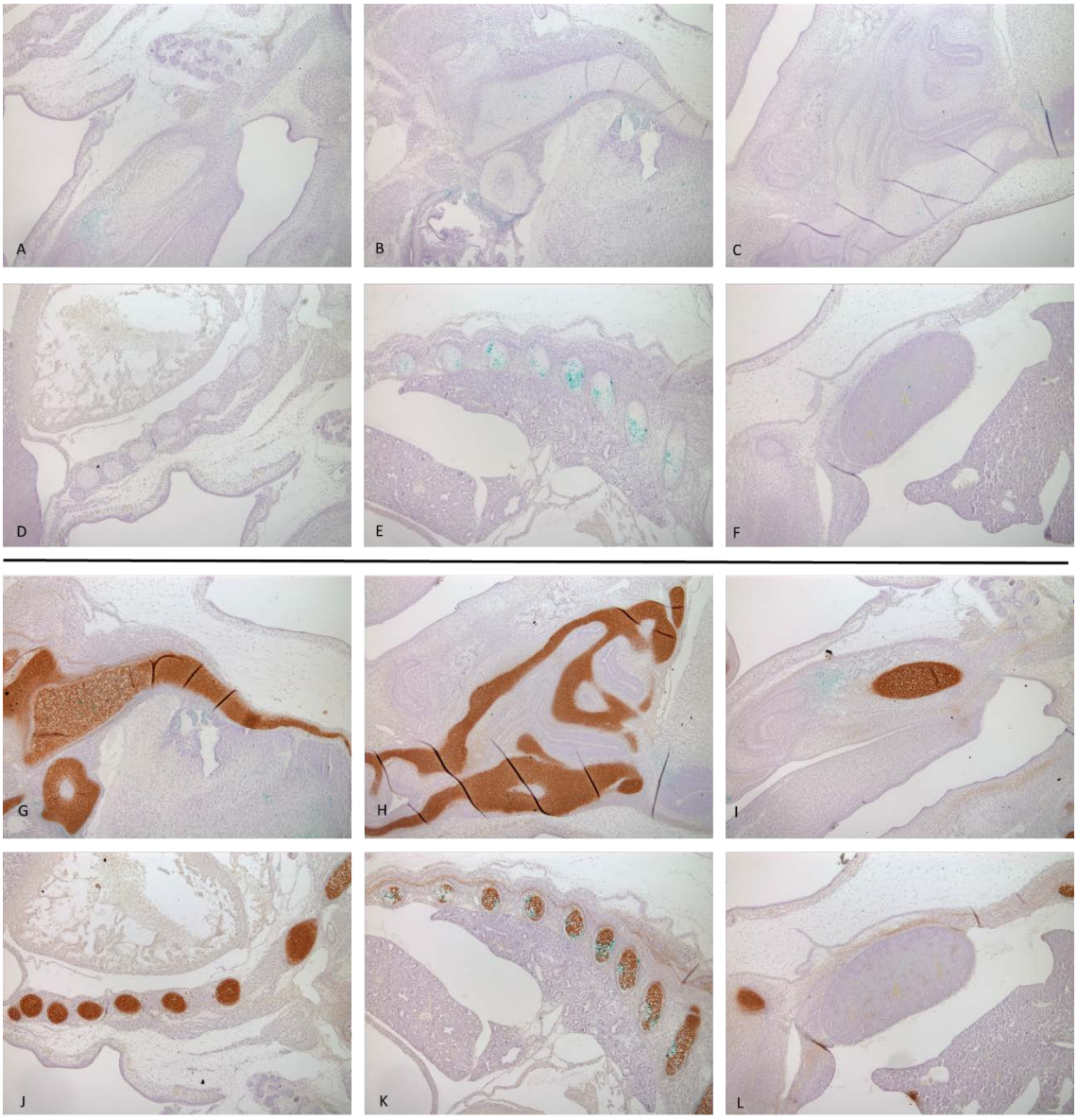
To determine if Acan expression co-localises with the +55kb enhancer staining sections were stained for the Acan G1 domain using the BR1 antibody.
To test the antibody sections from the -87kb enhancer embryos at E15.5 was stained. The antibody localises to Acan expression in cartilage elements such as the ribs, skull and Meckel's but not in the kidney at this stage (Figure 62).

In wildtype and +55kb enhancer kidneys examined at 8 weeks of age shows Acan localises to the tubules rather than the glomeruli (Figure 63). Antibody staining in the kidney was masked by the x-gal staining rendering co-localisation difficult (Figure 65). In the lungs Acan is found mainly in the vessels (Figure 64). Similar to the kidney the x-gal staining prevented the antibody binding in the +55kb lung (Figure 66). In the IVD of the +55kb enhancer analysed by BR1 staining the binding localises to the growth plate and the annulus fibrosis that surround the nucleus pulposus, which also expresses Acan but areas are masked by the x-gal staining (Figure 67). It is interesting to note that although the kidney and lung sections were from mice at the same age and processed the same, the antibody staining in the wildtype kidney tubules was not observed in the x-gal positive +55kb kidney. The reason for this is unclear, but repetition of this experiment would allow us to Reveal the extent of this staining.

The BR1 antibody rarely co-localises with the +55kb enhancer x-gal staining in the IVD and does not overlap in the kidney or lung suggesting that the enhancer may not drive expression of Acan in these tissues. The nearest gene to the +55kb other than *Acan* is *HAPLN3* therefore we examined the protein expression of *HAPLN3*, however, antibody

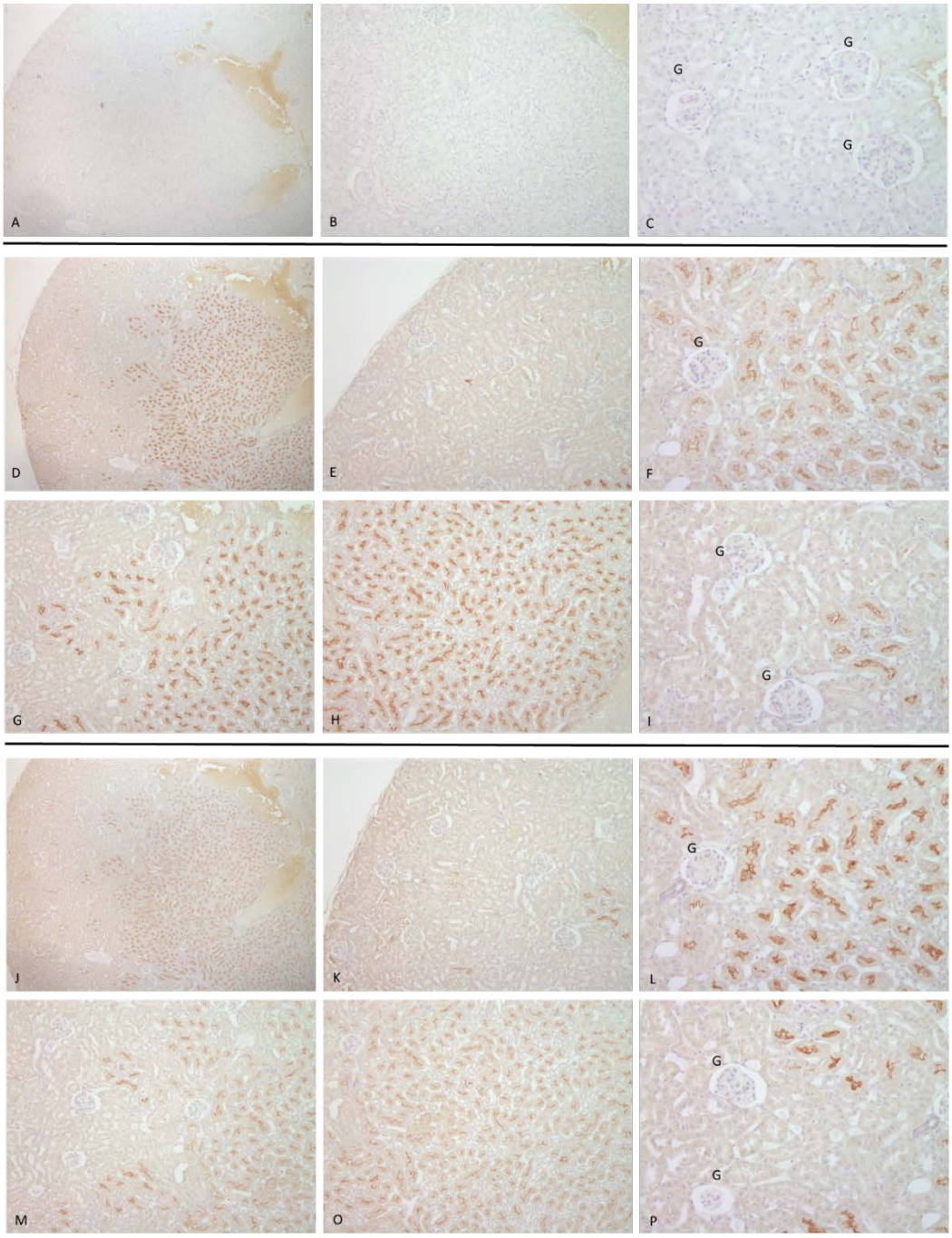
staining doesn't not localise HAPLN3 to the glomeruli but rather the medulla of the kidney and the bronchioles of the lungs (Figure 68 Figure 69).

Figure 62: BR1 antibody staining in E15.5 embryos



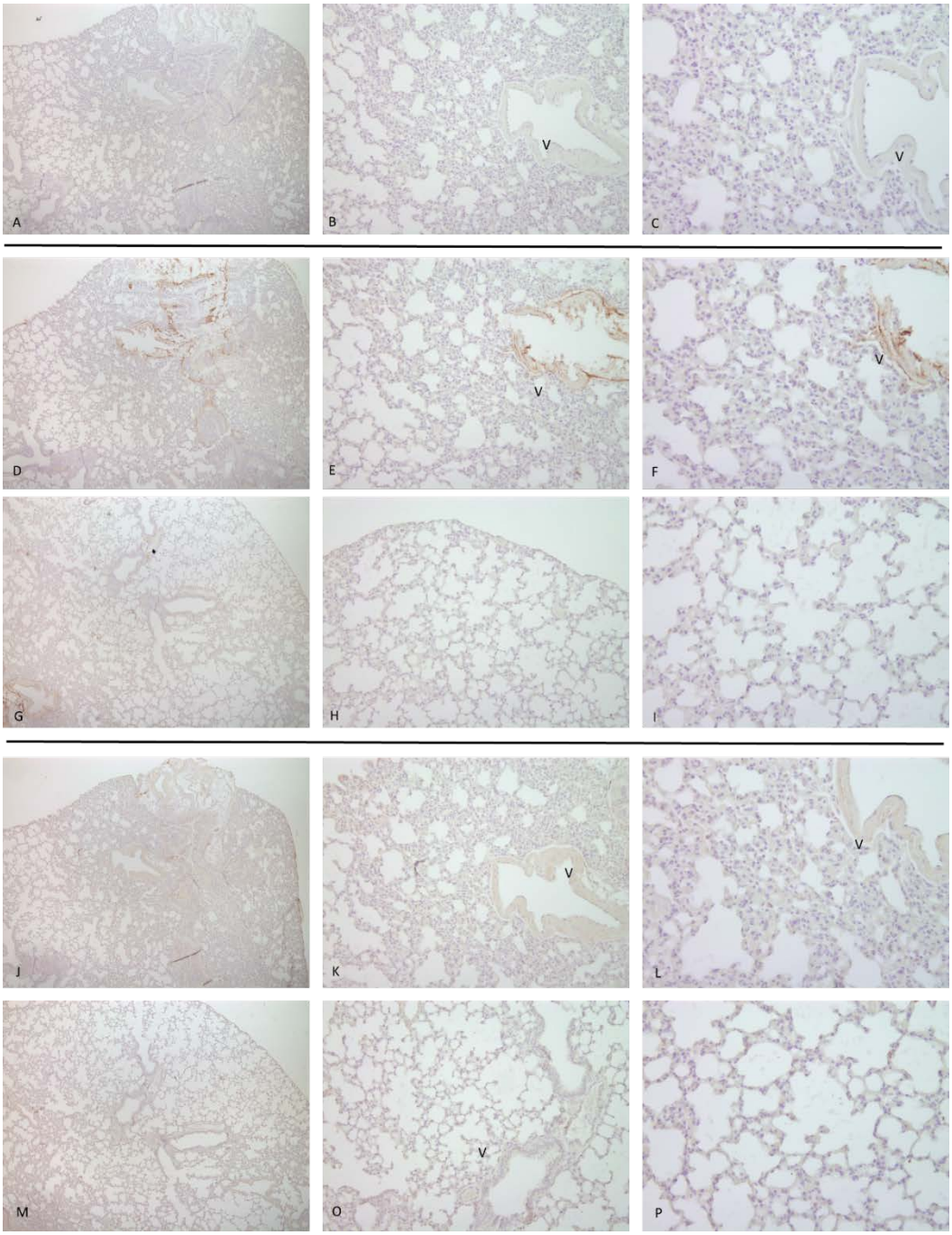
BR1 antibody staining on E15.5 embryo sections from the -87kb enhancer as this was the sections available at the time. A-F is no antibody control. G-L are the primary BR1 antibody. Staining is localised to the cartilage elements only in the skull (G and H), Meckel's (I) and ribs (J and K) but not in the kidney (L).

Figure 63: BR1 antibody in wildtype 8 week old kidney



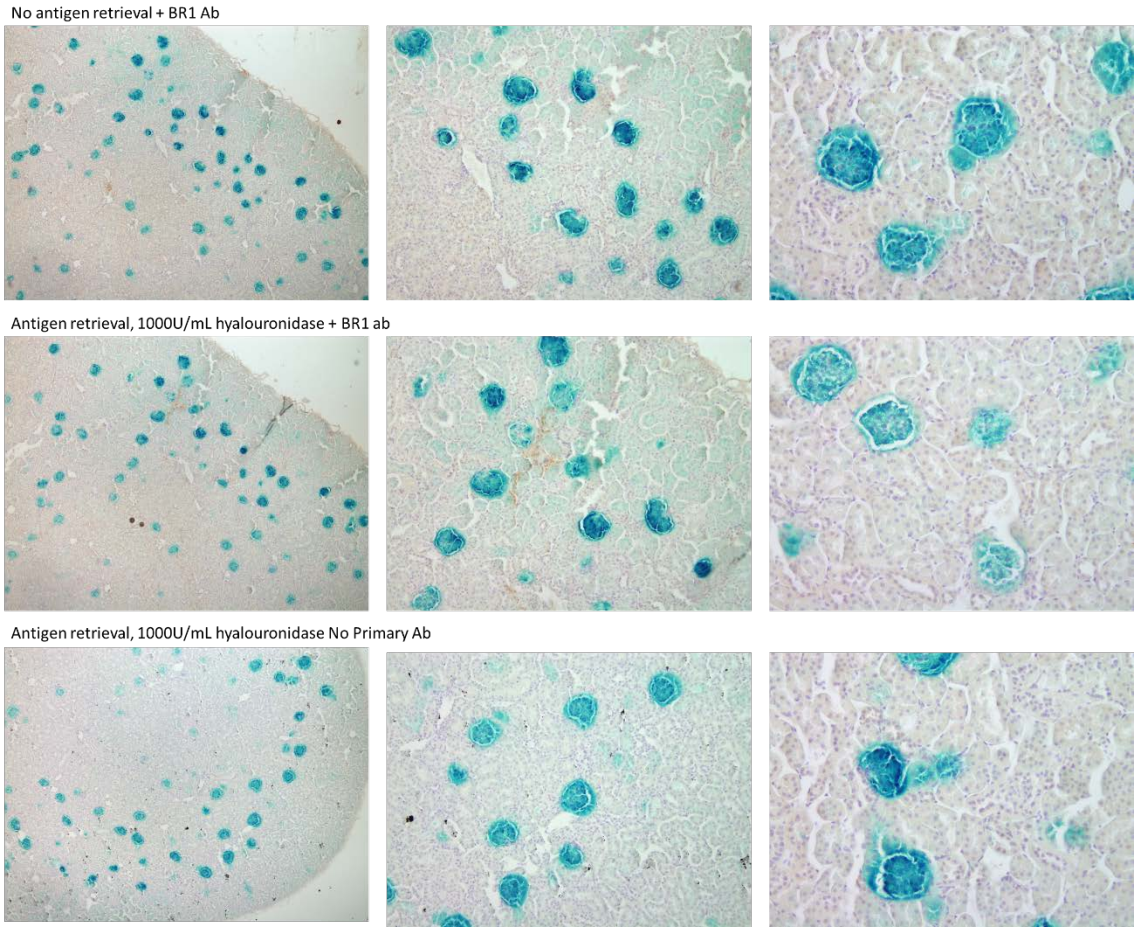
BR1 antibody staining on wildtype kidney. A-C are the no primary antibody control, D-I are the no antigen retrieval controls and J-P are the BR1 stained. Staining is localised to the tubules but not present in the glomeruli [G].

Figure 64: BR1 antibody staining in wildtype 8 week old lungs



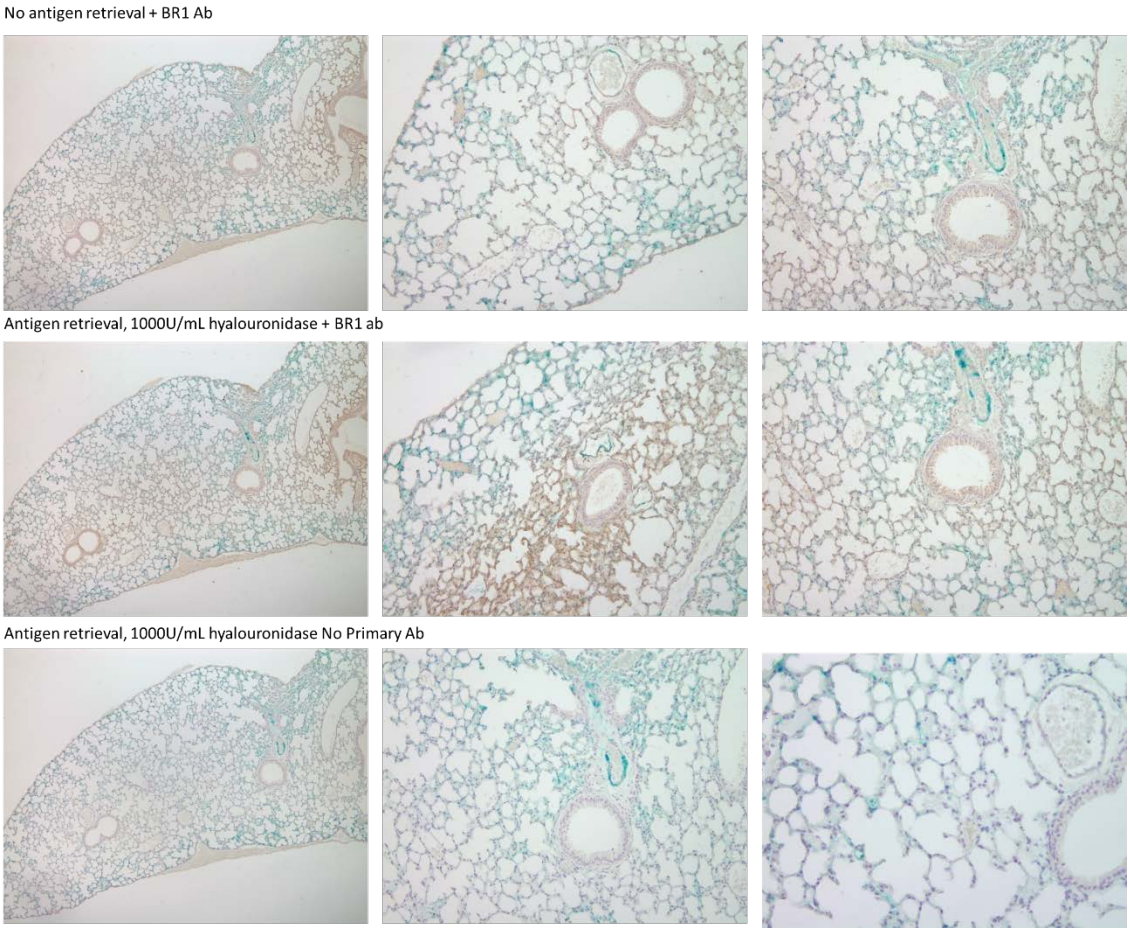
BR1 antibody staining on wildtype lung. A-C represent no primary antibody control, D-I are the no antigen retrieval controls and J-P are the BR1 stained. Staining is localised to the vessels [V] and weakly in the alveoli.

Figure 65: BR1 staining on the +55kb enhancer 8 week old kidney



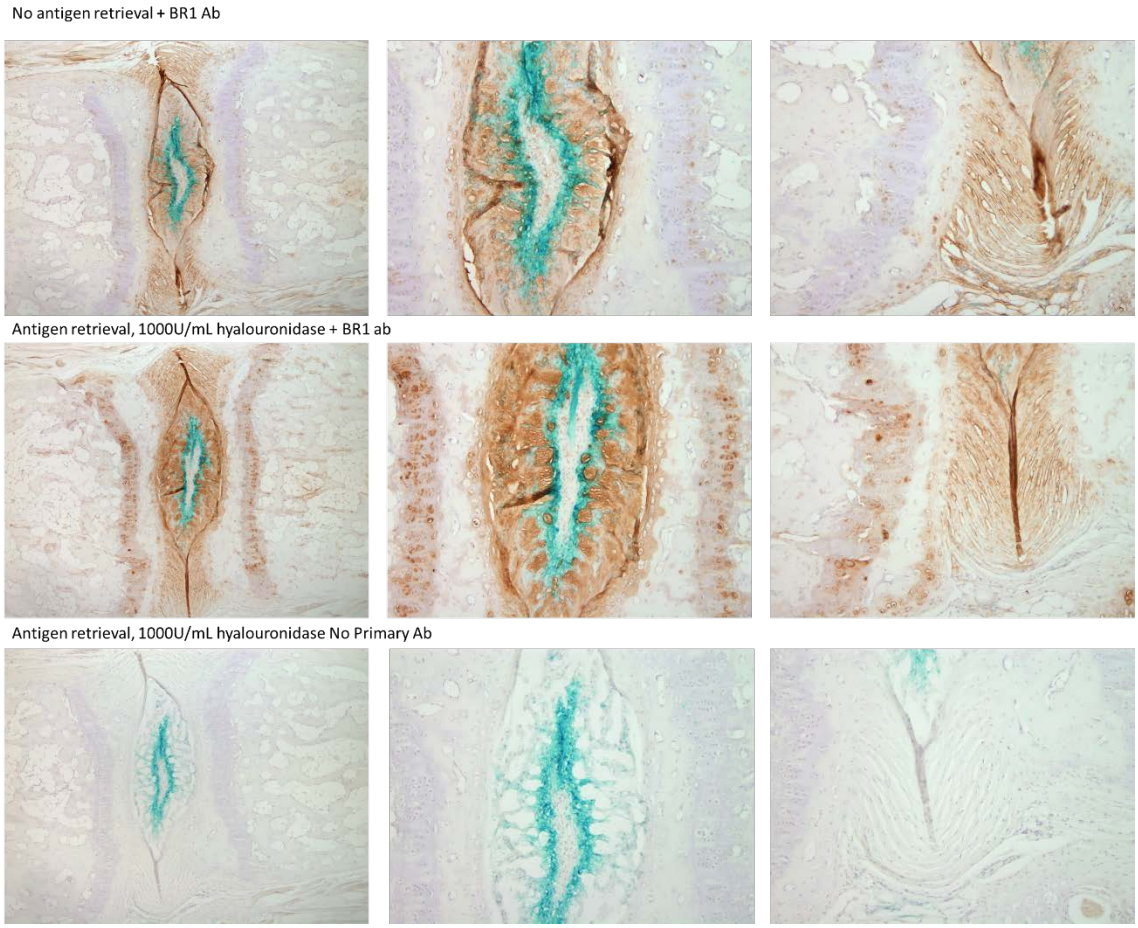
BR1 antibody staining on +55kb X-gal stained kidneys of 8 weeks of age. The X-gal staining appears to block the antibody binding.

Figure 66: BR1 staining on +55kb 8 week old lungs



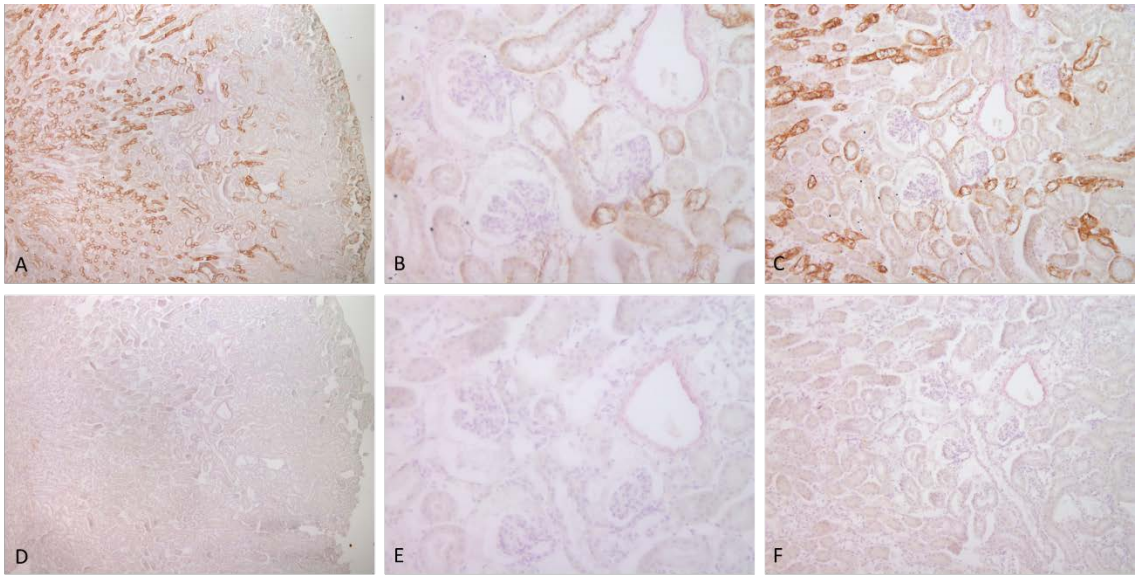
BR1 antibody staining on +55kb X-gal stained lung of 8 weeks of age. The X-gal staining appears to block the antibody binding. Staining is present in the vessels.

Figure 67: BR1 antibody staining on +55kb enhancer 8 week old IVD



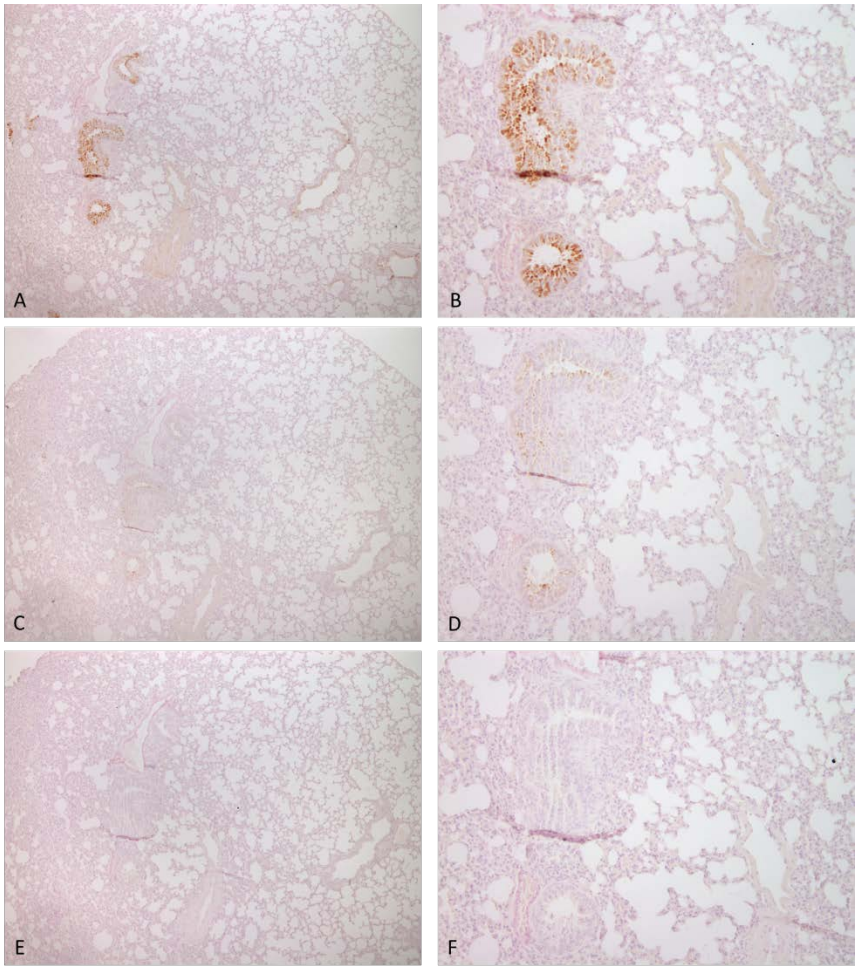
BR1 antibody staining on +55kb X-gal stained IVD of 8 weeks of age. The X-gal staining appears to block the antibody binding in the nucleus pulposus. Staining is observed in the chondrocytes of the growth plate, in the annulus fibrosis and in the nucleus pulposus but not where the X-gal staining is present.

Figure 68: HAPLN3 antibody staining in 8 week old mice kidney



8 week old wildtype kidney stained with primary antibody raised against HAPLN3 with antigen retrieval (A-C) and no primary antibody (D-F). Staining resides mainly in the medulla (A) and does not stain the glomeruli (B).

Figure 69: HAPLN3 antibody staining in 8 week old mice lungs



8 week old wildtype lungs stained with primary antibody raised against HAPLN3 with antigen retrieval (A and B), no antigen retrieval (C and D) and no primary antibody (E and F). Staining is localised to just the bronchioles and no other structures (B).

Table 10: Overview of expression pattern of the six *Acan* enhancers

Tissue	Intensity of staining					
	+60kb	+55kb	+26kb	-35kb	-65kb	-87kb
Chondrocranium	±	-	+++++	+++++	++	±
Nasal cartilage/ olfactory	±	-	++++	+++++	-	+
Meckle's cartilage/ mandible	-	-	+++++	+++++	-	-
Costal cartilage of the ribs	±±±	-	++++	+++++	+++	+
Femur	±±	-	++++	+++++	+++	++
Tibia	±	-	++++	+++++	+++	++
Fibula	±±	-	++++	+++++	+++	++
Digits of hindlimb	-	-	++++	+++++	+	+++
Digits of forelimb	-	-	++++	+++++	+	+++
Scapula/ humerus	±	-	+++++	+++++	+++	++
Ulna	±	-	+++++	+++++	+++	++
Radius	±	-	+++++	+++++	+++	++
Intervertebral discs	-	++++	+++++	+++++	-	-
Lumbar spine	-	++	+++++	+++++	++	++++
Cartilage of the ear/ occipital	-	-	+++++	+++++	+++	++
Tail	±	++++	+++++	+++++	+	-
Brain	-	-	++	±	-	-
Vibrissae	-	-	+++	+++	-	±
Heart	-	-	-	-	-	-
Lung	-	+	-	-	-	-

Liver	-	-	-	-	-	-
Tongue	±	-	-	-	-	-
Skin	±	+	-	-	+++	-
Kidney	±	+++	+++	-	-	-
Blood vessels	±	+++	-	-	-	-

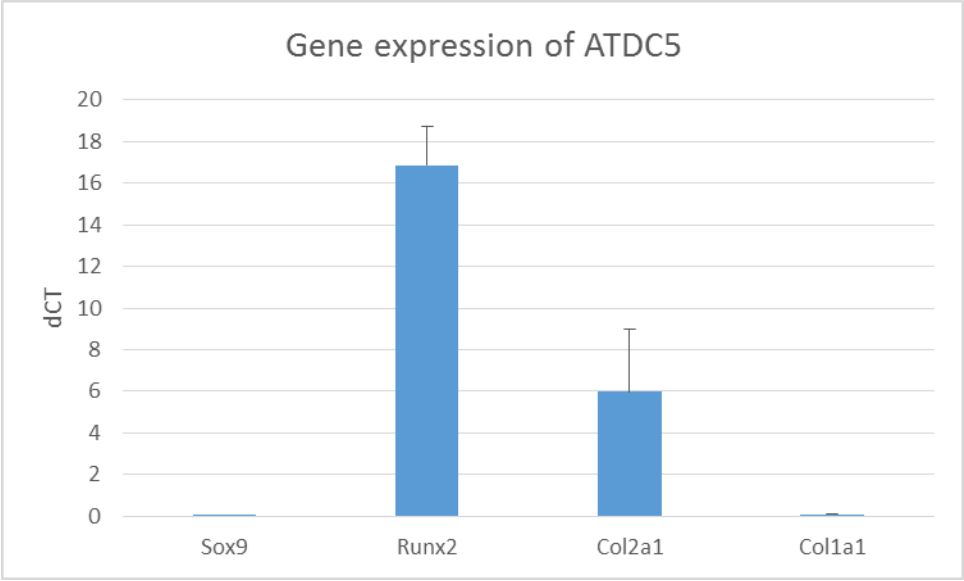
Summary of the staining intensities and location for the Acan enhancer. – is not detected; + weakly detected; ++ not completely uniform staining throughout tissue; +++ majority of tissue stained, ++++ strong staining in entire tissue; +++++ entire tissue is stained and has intense signal; ± present in some founder embryos but not all.

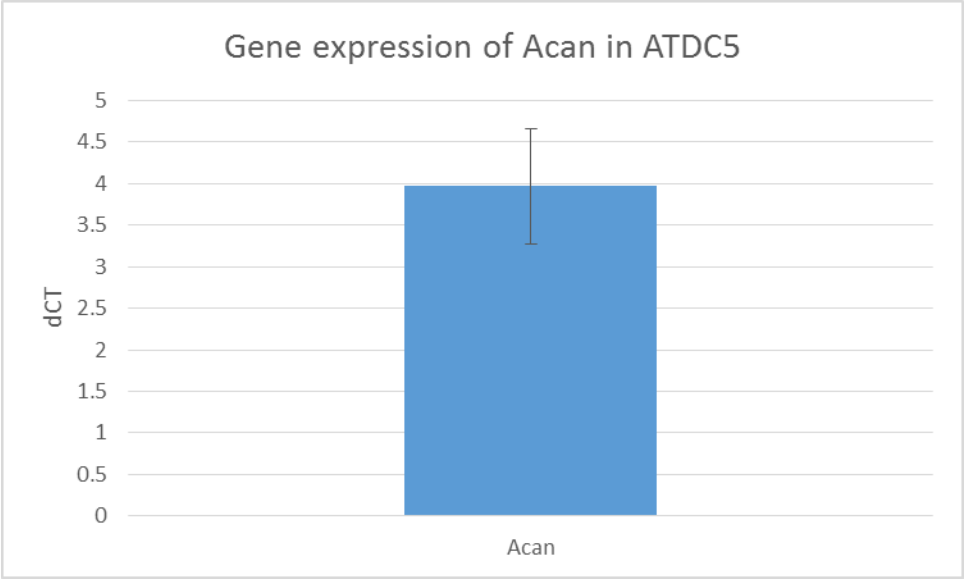
4.4. Analysis of the Acan enhancers in vitro

4.4.1. Characterisation of ATDC5 expression of cartilage genes

The ATDC5s had been differentiated for 20 days and to determine the gene expression qPCR was undertaken. The ATDC5 expressed high levels of Col2a1 with little Col1a1 suggesting a chondrocyte phenotype. However, the Sox9 content was low, which was concerning for chondrocytes, whereas the Runx2 expression was relatively higher in comparison to the Col2a1 and Sox9. In terms of aggrecan expression there is detectable levels of *Acan* and lower levels of TIMP3, this allows us to perform further assays using the ATDC5 cells.

Figure 70: Relative gene expression of ATDC5 differentiated for 20 days.



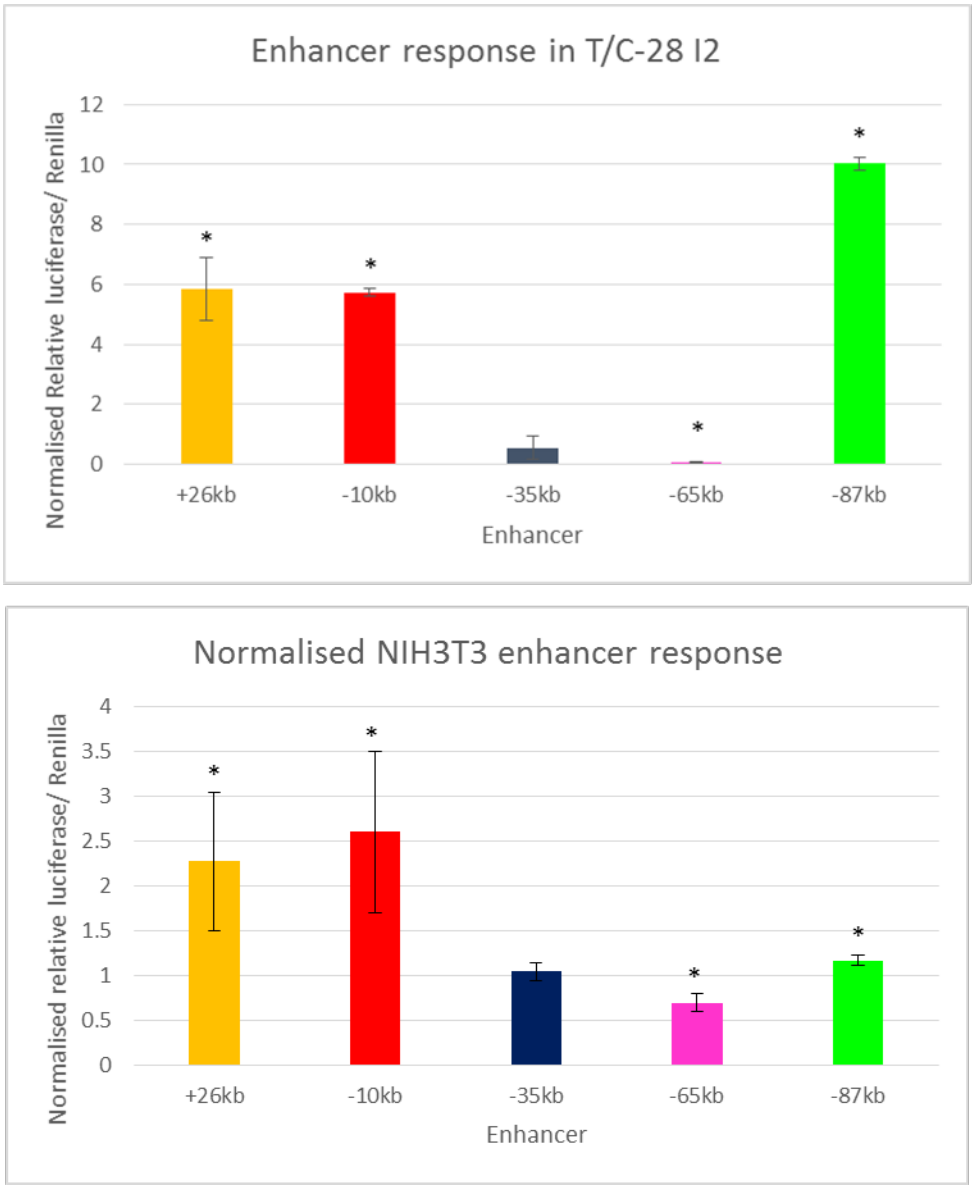


qPCR results from ATDC5 cells when RNA is extracted after 20 days differentiated. The ATDC5 resemble chondrocytes by expressing more Col2a1 than Col1a1 but there are lower levels of Sox9 with more Runx2. Additionally, Acan is expressed in the ATDC5 detected by Taqman probes. Error bars are expressed as SEM.

4.4.2. Expression of *Acan* enhancers *in vitro*

pGl3 promoter vectors with the *Acan* enhancers cloned upstream were examined for activity in T/C28-I2 and NIH3T3 cells initially. The -35kb, unlike *in vivo,* expresses weakly in the chondrocyte cell line as well as the -65kb enhancer. The +26kb, -10kb and -87kb expresses highly. In the fibroblast cells the +26kb and -10kb enhancers retain high expression when compared to vector alone whereas the -87kb is reduced suggesting a chondrocyte affinity of the -87kb. Additionally, the -65kb and -35kb retain a low level of activity in the fibroblast.

Figure 71: Normalised response of enhancers in chondrocytes and fibroblasts *in vitro*

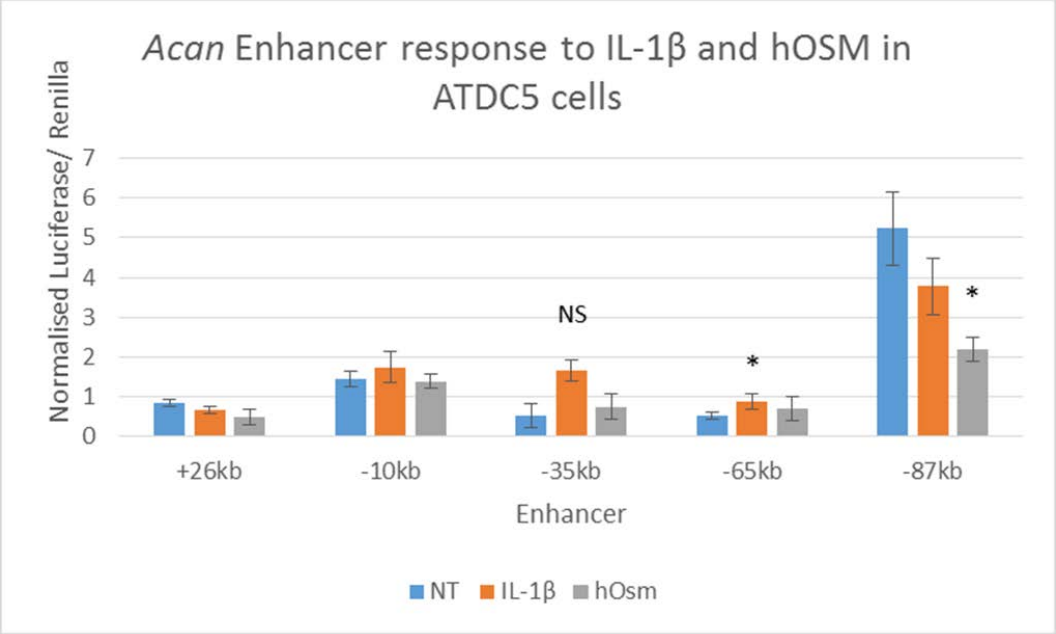


*In T/C28-I2 cells the enhancers +26kb, -10kb and -87kb have high level of activity whereas the -35kb and -65kb expresses weakly compared to vector control. In comparison in the NIH3T3 the -87kb has reduced expression but the +26kb and -10kb have relatively higher expression compared to vector alone but still low levels of -35kb and -65kb activity. *p=<0.05, Error bars are SEM.*

4.4.3. Response of *Acan* enhancers *in vitro* to catabolic agents in ATDC5 cells

The ATDC5 cells were used to examine the effects of hOSM and IL-1β on the activity of the *Acan* enhancers. Upon transfection and then treatment with the cytokines the cells were assayed for luciferase activity. There is no significant effects of these agents on the +26kb and -10kb enhancer. There is a slightly non-significant increase in the -35kb enhancer to IL-1β treatment. The -65kb increases in activity significantly in the presence of IL-1β and the -87kb enhancer's activity is dampened with hOSM treatment.

Figure 72: Effects of catabolic agents on enhancers' response in ATDC5



*ATDC5 cells are transfected with luciferase reporter vectors for the chondrocyte specific Acan enhancers and then treated with either human oncostatin M (hOSM) or IL-1β and then assayed for luciferase activity after 24hours. The effects of the two agents have no significant effects on the +26kb, -10kb and -35kb enhancers. The -65kb responds to IL-1β treatment by increasing expression whereas hOSM treatment reduces the activity of the -87kb enhancer. *p=<0.05 and error bars are presented as SEM.*

4.5. Analysis of -35kb Acan Cre

4.5.1. Background

As previously mentioned, enhancers are capable of driving tissue specific expression of a target gene. This quality allows the enhancer sequence to be used to analyse temporal and spatial effects of knockouts and knock-ins by driving Causes recombination (Cre) enzyme in a tissue specific way while avoiding some of the lethality of embryonic stem cells (ES cells) knock-outs (Heffner et al., 2012, Sharma and Zhu, 2014). The Cre recombinase is a Bacteriophage P1 38 kDa protein that catalyses intra and inter molecular recombination of DNA in a site specific way. The recombination occurs between a 34bp sequence known as *loxP* with a minimum of 82bp between each site to be efficient and is direction dependant (Sauer and Henderson, 1988).The *loxP* sites can be in the same direction on the chromosome, different chromosome or in opposite directs to result in deletions, duplications, translocation and inversions. The Cre can be engineered to only act or activated in inducible manner by using tetracycline or by using drugs. A fusion of a triple mutated human ligand binding domain (G400V/M543A/L544A) of the oestrogen reception allows the enzyme to be attached to HSP90 in the cytoplasm therefore not active. Upon injection of 17-β-estradiol analogous such as tamoxifen acts as an antagonist of the receptor, allows the translocation of the Cre into the nucleus and allows time dependant activation. The first oestrogen responsive Cre is known as CreER and a highly sensitive version has been engineered to be 10 times more reactive once tamoxifen is injected ($CreER^{T2}$) (Indra et al., 1999, Brocard et al., 1997).

There has been many Cre's designed to target chondrocyte progenitor cells, Prx1-Cre (Logan et al., 2002),GDF5- cre (Rountree et al., 2004) , Sox9 (Akiyama et al., 2005) and others targeting differentiated chondrocytes such as Col10, Col11 and Col2-cre (Ovchinnikov et al., 2000, Hilton et al., 2007, Koyama et al., 2008, Fujimaki et al., 2005).Unlike *Col2a1*, *Acan* is expressed continuously in chondrocytes and the -35kb enhancer does not have other non-chondrogenic expression seen in the *Col2a1*-cre such

as in the kidney (Kolpakova-Hart et al., 2008, Long et al., 2001, Fosang et al., 2013). Additionally, there is a current dual luciferase and Cre that uses the -10kb enhancer in tandem repeat that is able to drive expression in chondrocytes following the expression of *Acan* and allows real time tracking of the transgene (Cascio et al., 2014). Another *Acan* –cre was generated by knocking in the IRES-Cre-ER^{T2} downstream of the endogenous *Acan* stop codon that drives expression even at 6 months (Henry et al., 2009). This cre allows the analysis of developing and adult cartilage which is beneficial to studying age related diseases such as OA. Therefore, we generated an *Acan*-Cre using the -35kb and the human *Acan* minimal promoter to drive the CreER^{T2} named pChu9 and generated 4 lines for interrogation.

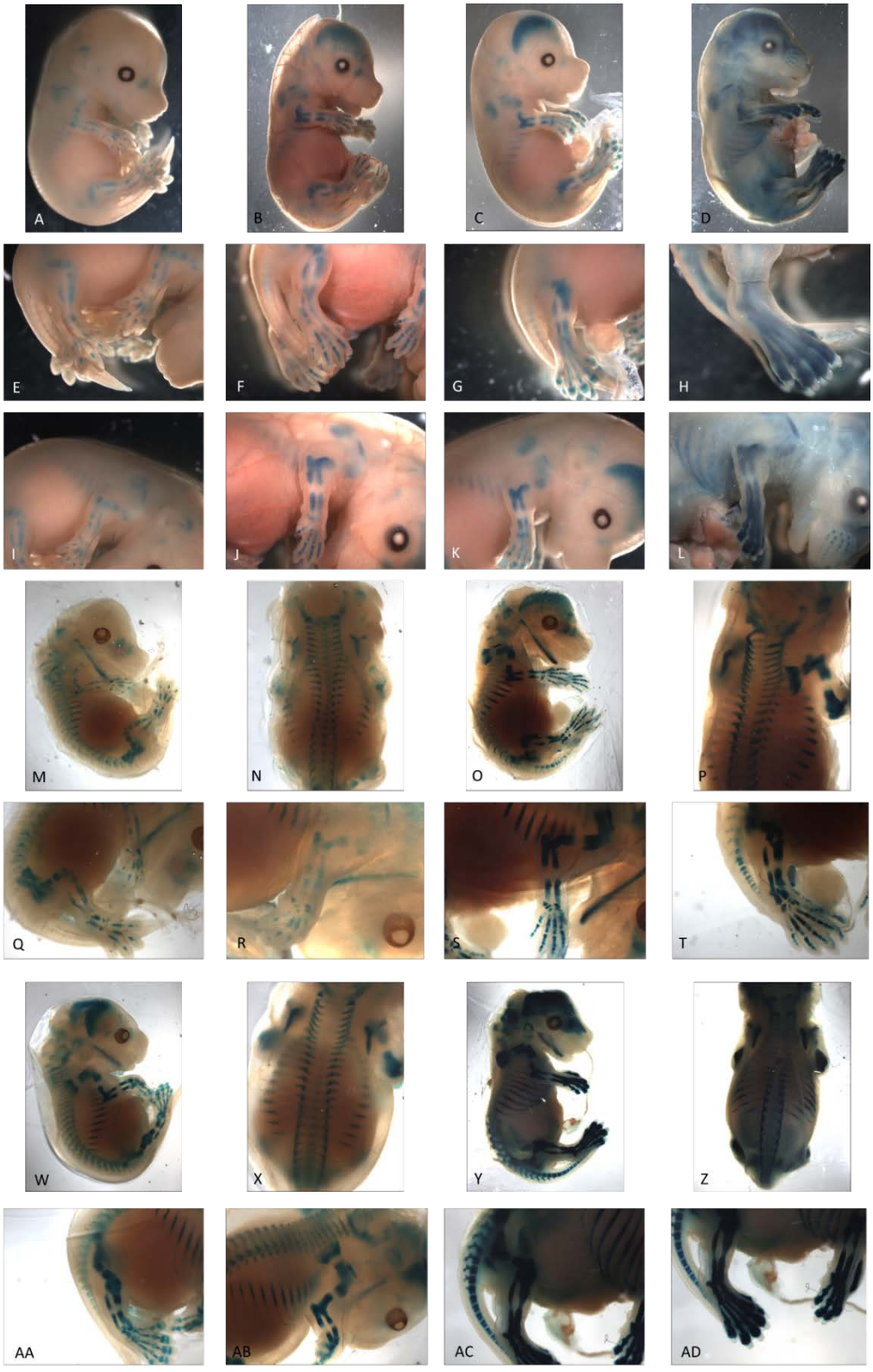
To examine the response and activity of the Cre we crossed the pChu9 mice with R26R reporter mice that contains a floxed *LacZ* visualise cells that the Cre is active in during development (E15.5) and adult mice (Soriano, 1999).

4.5.2. Determination of the expression of the pChu9 lines

To establish the expression pattern of the pChu9 lines and to ensure the -35kb enhancer is reproduced under the influence of the human promoter. Female pChu91, pChu92, pChu93 and pChu94 mice were time mated with male R26R overnight, after checking for successful copulation plugs this day was classed as E0.5. On day E13.5 tamoxifen was injected and embryos were collected at E15.5 and stained for β-gal activity. Whole mount showed variability in the strength of expression in the craniofacial, mainly the skull. pChu91 showed little or no expression in these areas but still expressed in the skeletal elements seen in the -35kb enhancer, but was the weakest of the 4 lines (Figure 73A, E and I). pChu92 (Figure 73B, F and J), pChu93 (Figure 73C, G and K) and pChu94 (Figure 73D, H and L) showed similar expression in the limbs and ribs. However, there were differences seen in the tail of the lines with pChu94 expressing the strongest (Figure 73H). Clearing the embryos with 1% KOH and glycerol reveals the skeletal element staining more clearly, pChu91 (Figure 73 M, N, Q and R), pChu92 (Figure 73W, X, AA and AB), pChu93 (Figure 73O, P, S and T) and pChu94 (Figure 73Y, Z, AC and AD) all

show good staining in the limbs, spine, jaw, ribs, metacarpals and tarsals but pChu94 shows more intense spinal and tail expression (Figure 73Z and AC).

Figure 73: Whole mount images of E15.5 pChu9 lines



E15.5 whole mount images of pChu9 lines and cleared embryos shows skeletal element staining in the spine, ribs, fore and hind limb. Each line expresses pChu91 (A, E and I), pChu92 (B, F and J), pChu93 (C, G and K) and pChu94 (D, H and L) to varying degrees, cleared embryos show the expression pattern clearer, pChu91 sagittal (M), back (N), hindlimb and tail (Q) and the forelimb (R). pChu92 sagittal (W), back (X), hindlimb and tail (AA) and the forelimb (AB). pChu93 sagittal (O), back (P), hindlimb and tail (T) and the forelimb (S). pChu94 sagittal (Y), back (Z), hindlimb and tail (AC) and the forelimb and hindlimb (AD).

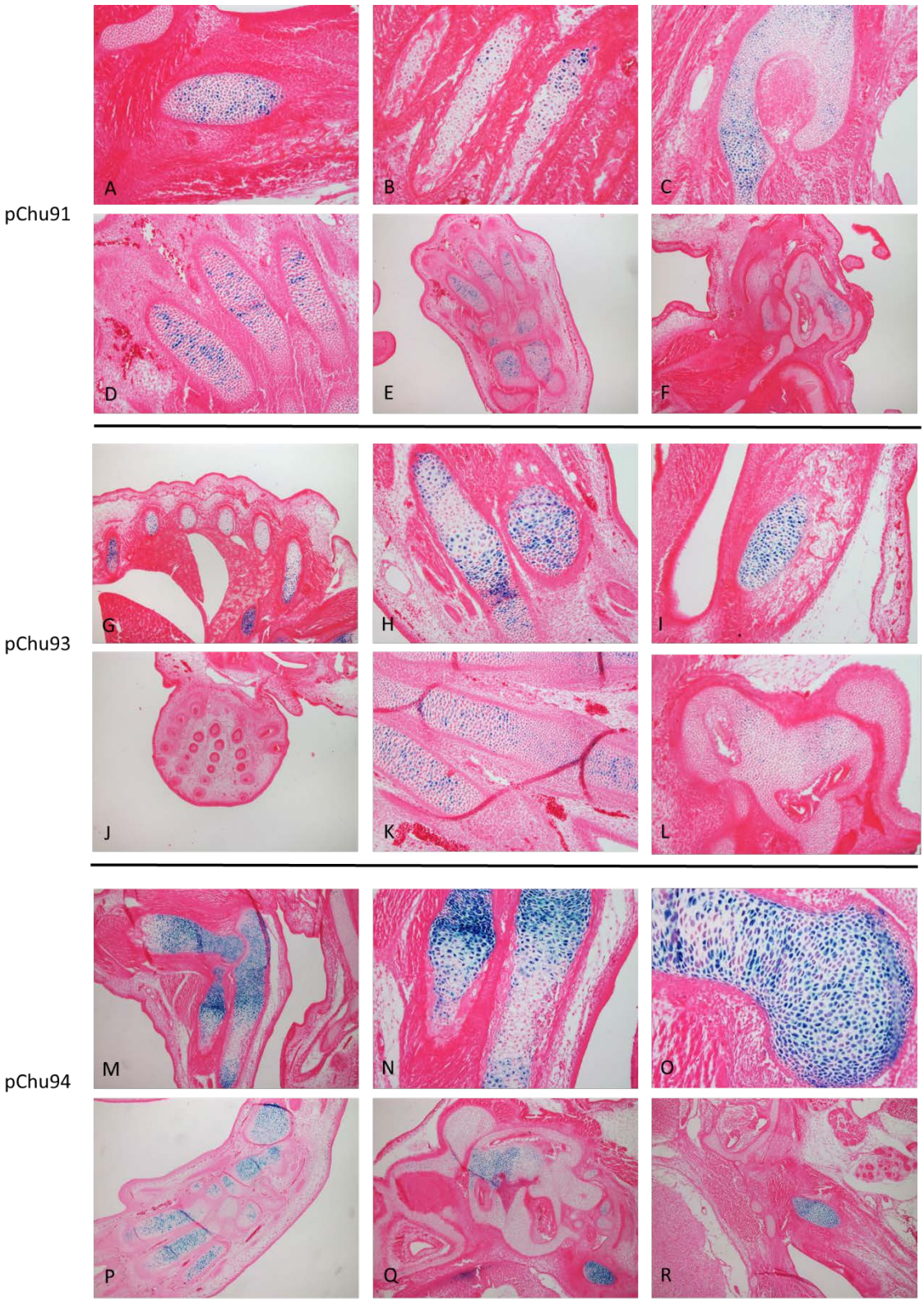
Histological examination of the pChu91, pChu93 and pChu94 show that they express in skeletal elements. The pChu91 shows scattered expression in Meckel's (Figure 74A), ribs (B), femur (Figure 74C), metacarpals and carpels (Figure 74D and E) but little Figure 74expression in the ear (Figure 74F), the expression is chondrocyte specific. pChu93 shows more intense staining than pChu91, but still scattered in the ribs (Figure 74G), radius, ulna (Figure 74H) and metacarpals (Figure 74K) Staining is stronger in Meckel's cartilage (Figure 74I) but little is detected in the ear (Figure 74L) and absent from the vibrissae (Figure 74J) unlike the -35kb enhancer driving the HSP68 minimum promoter. The pChu94 mimics the pChu93 in expressing strongly in the paw (Figure 74P) and Meckel's (Figure 74R), however, more staining is observed in the ear (Figure 74Q). All the chondrocytes in the pChu94 have had the cre active with staining seen in the humerus, radius and ulna (Figure 74M, N and O).

4.5.3. Activation of the pChu9 lines at 8 weeks shows expression in cartilage elements

To determine if the pChu9 lines were active in adult mice, injections of tamoxifen was carried out at 8 weeks of age and the mice culled and tissue collected after 8 weeks. In whole mount tissue the no tamoxifen pChu94 showed no expression in any of the tissue (pChu94-Tx), whereas there is strong activity in the pChu94 (pChu94 +Tx) with lower intensity in pChu93 (pChu93 +Tx) and the transgene is poorly activated in the pChu92 (pChu92 +Tx). There is strong staining in the tail, knee, superficial region of the scapula, femoral head and in the ribs of the pChu94 when the Cre is activated compared to the

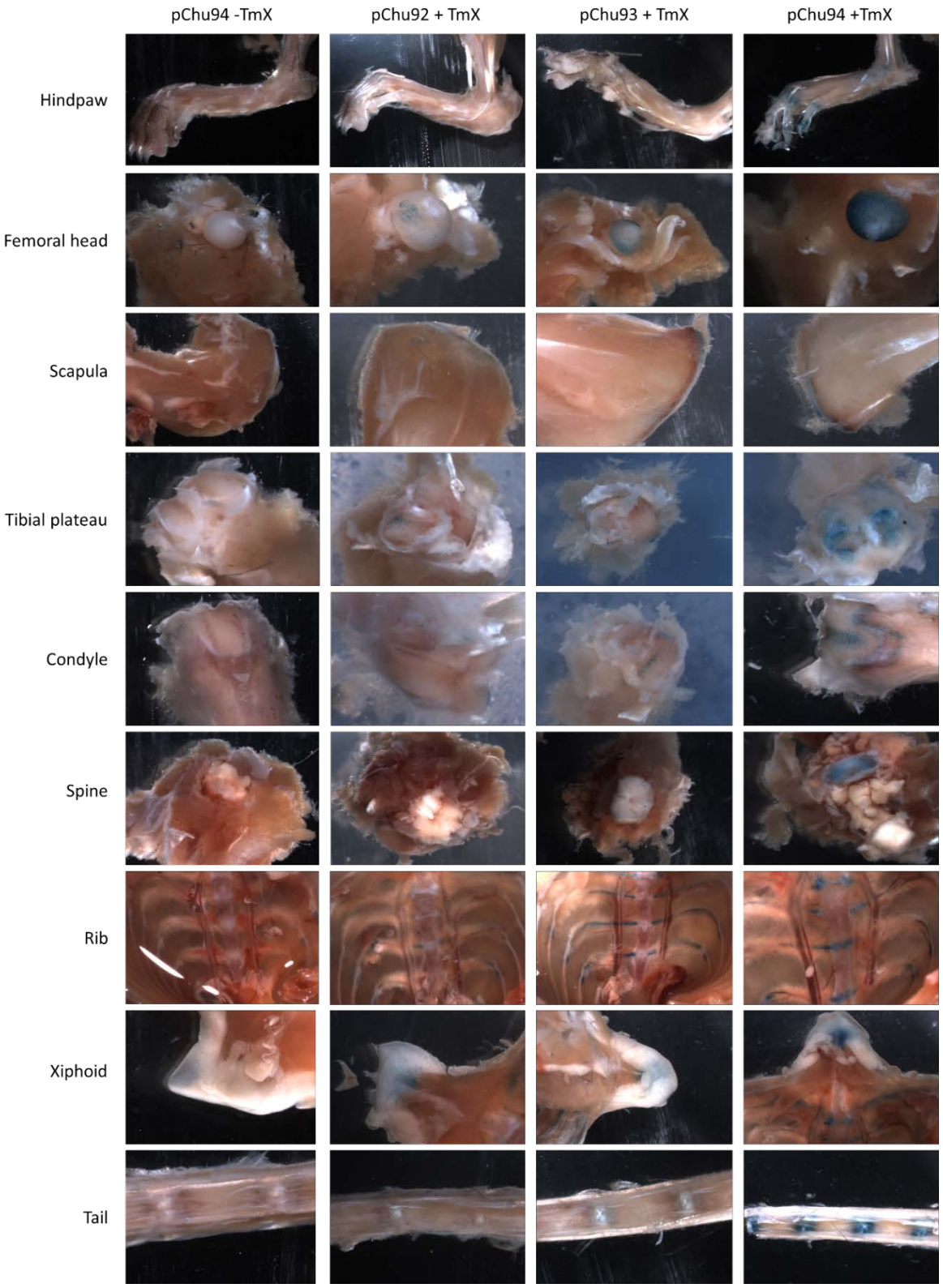
other lines. There is also spinal cord staining not seen in the other lines. Unfortunately due to time constraints histological examination of the tissue was not performed. With the lack of histological examination of the adult tissue we cannot say for certain if the cre recombinase is efficient in removing the floxed gene in adult tissue, this may be a hindrance to any future use of this transgenic mouse line.

Figure 74: Histological examination of the pChu9 lines at E15.5



Sections of pChu91, 3 and 4 counter stained with eosin. pChu91 expresses in chondrocytes but the activation of the Cre is scattered and weak in terms of intensity in the Meckel's cartilage (A), ribs (B), femoral head (F), metacarpals and fore paw (D and E) and ear (F). The pChu93 activation is greater than the pChu91 but still scattered in the ribs (G) and metacarpals (N). However, in Meckel's (I) and in the radius and ulna there is consistent activation but it is not uniform (H). There is little expression in the ear (L) and no expression in the vibrissae. The pChu94 expresses the strongest compared to the other two lines in the limb (M) seen in higher magnification of the radius and ulna (N) and humerus (O) as well as Meckel's (R)and the forepaw (P). The ear is weaker and non-uniform in expression (Q), which is different to the -35kb enhancer.

Figure 75: Whole mount images of adult tissue of the pChu9 lines



Adult tissue from the pChu92, 3 and 4 lines injected with tamoxifen at 8 weeks of age and collected 8 weeks after shows that the pChu94 Cre is activated more successfully than the other two when compared to no tamoxifen although the spinal cord is also staining for β-gal activity. TmX: tamoxifen.

5. DISCUSSION

5.1. The use of in vivo mouse models versus in vitro cell assays for assessing cis-acting elements

In this study transient and stable mouse lines were analysed to characterise *cis*-acting elements of *Acan* this was done prior to analysis within immortalised cells lines. The reasoning to complement *in vivo* with *in vitro* was due to the nature of studying enhancers. *In vitro* studies do not show appreciation of interactions between different cells or understanding how the enhancer is working temporally or spatially in deed there is no appreciation of how enhancers express in the whole organism. Additionally, the use of cells is a more targeted towards the available cell lines and primary chondrocytes are difficult to obtain from adults and perform poorly in tissue culture conditions due to the loss of ability to divide. Furthermore, they may not recapitulate the expression that is seen *in vivo* as popular immortalised human rib chondrocytes T/C-28- I2, T/C28-A2 and T/C 28-A4 express lower matrix genes, show reduced synthesis and turnover of matrix protein therefore variability in cell culture may not recapitulate the expression of ACAN seen *in vivo* (Finger et al., 2003). Furthermore, there are limited similarities between SW1353 (human chondrosarcoma cells) and primary human chondrocytes with limited response to cytokine treatment (Gebauer et al., 2005). In contrast, the chondrocyte within tissue is in a high load and hypoxic conditions, normal tissue culture conditions are atmospheric oxygen (around 21%) in a static condition therefore not physiological. On the other hand, the use of cell lines cannot be completely discredited, the mouse cells ATDC5, that was also used in this study are embryonic carcinoma chondroprogenitors clones that can be induced through an insulin-dependent pathway to follow growth plate cartilage development *in vitro,* they undergo cellular condensation after 5 days of treatment and cartilaginous nodules form at 11 days up to 31 days can be used at different stages to recapitulate *in vivo* development (Suwanwela et al., 2011, Shukunami et al., 1997, Atsumi et al., 1990). As ATDC5s undergo induction they express collagen type 2 and 5 during early phases and collagen 9 and 11 at the later

stages with collagen 10 being the last collagen to start expressing as cells undergo hypertrophy Other ECM genes such as link protein and *Acan* are expressed very well by these cells once differentiation occurs, however, hypoxia delayed early chondrogenesis and suppressed hypertrophy which underlines differences between normal physiology and cultured conditions but the relative effects of an enhancer, especially for *Acan* expression can be examined *in vitro* with degrees of confidence (Chen et al., 2005, Shinomura et al., 2006, Chen et al., 2006). Furthermore, in a collaboration with the University of Newcastle, EMSAs were performed using nuclear protein extracts (with the same extraction protocol) from the human chondrosarcoma cell line SW1353 from cells cultured at Newcastle University and at the University of Liverpool, UK, were examined for Sox9 binding and compared to the IVT protein. The same cell line cultured in different labs contained different binding patterns for Sox9 but neither cell line contained the correct band size for Sox9 and super-shifts only showed shifts in the IVT proteins (Figure 85). Cells can, depending on the lab that has prepared and cultured, have phenotypic, genotypic and karyotypic drifts. This can result from prolonged culture times, high passage number, cell senescence and sub-population selection (Herrmann et al., 2007). Population selection can arise from faster growing cells, sheer forces from pipetting or even trypsinisation times, which means with time the population of cells will diverge slowly from the original culture. This shows the variability within cultured cells and a minor disadvantage of using certain cell lines for enhancer examination.

In vivo models are not without problems, difficulties in generating founders of lines plagued this project, with no detectable founders verified by genotyping after oocyte injections, mice were weaned and reach sexual maturity but time mating's with wildtype F1 females yielded no expression at E15.5 and genotyping of the embryos by tail snip showed no transgene, or transgene, but no expression. From females and confirmed males founders (F0) that expressed through time mating experiments then fostered other problems such as no germline transmission, as Mendelian ratio the offspring (F1) should in theory carry 50% of the transgene. This was not the case as in some cases there were no transmission, although F0's were positive, or the females

were genotype positive when the line was culled and the founder was taken for examination. This may mean that the F0's were mosaic that may have arisen from late integration of the transgene. Two other possibilities are that the transgene were silenced or the transgene was mutated. Due to the nature of transgenic mice experiments the transgene is inserted randomly and copy number can vary, in terms of gene silencing it is not unlikely that the cassette inserted into a transcriptionally repressed region and was unable to change the chromatin landscape to activate its own transcription regions such as repressors, DNA methylation sites, X-linked inactivation or genome imprinting, although rare can occur.

One other consideration for this project that must be highlighted, that are often over looked, are the environmental conditions the mice are subjected too. As mentioned previously, the animal unit at the University of Liverpool is SPF, this means that the mice are free from a wide range of potentially pathogenic bacteria, virus and parasites that are found in non-SPF condition. This has been shown to have effects on drug response and immune response. For example SPF housed mice have immune systems resembling that of new-born humans and changing the condition to housing shared with wild and store brought mice changed the immune system to one resembling that of adult humans (Beura et al., 2016). This has profound effects on the immune system response, mice when injected with cyclophosphamide, a drug that dampens the immune system and fragmented dsDNA, results in multiple organ failure from systemic inflammation and sepsis and eventually death in non-SPF housed animals, however, SPF mice do not display this phenotype (Dolgova et al., 2015). Another example is that in the AD10- anti-NGF mice have progressive neurodegeneration and cognitive decline but in a SPF-unit the behaviour of the mice are milder (Tzanoulinou et al., 2014). Although, he effects on transgenic visualisation cassettes and other transgenic mice have not been study extensively and may not have a major effect on the genetics of the mice. The remaining factor from the housing unit that credits a mention is the effects on human interactions on the mice, as noise is one factor that is not controlled and results is increased physiological distress in mice, although the genetic bases may be minimal the effects on

breeding could have had profound effects on the speed in which lines were analysed in this study (Lauer et al., 2009). During the time of study the animal unit has had two construction sites within close proximity that has caused noticeable vibrations and sound pressures into the unit detected by users and would have been identified by mice who have hearing that extends to the ultrasonic frequencies not heard by humans. It has been shown that excessive noise by construction equipment leads to an increase in still births, reduced litter sizes and reduced weight (Rasmussen et al., 2009, Reynolds et al., 2010).

In regards to the SOI and position effects of using transgene and the effects of inserting cassettes into the genome are considerations for the physiological relevance of these enhancer regions. However, enhancer can act irrespective of their location, the p53 responsive element controls stress response by contacting with the centromere when the wild type location is abolished and the sequence is placed in an ectopic locations in the genome the element is able to reconnect with the target gene (Link et al., 2013). This suggests that in this study the enhancers identified, should be able to regulate expression irrespective of their location but the effects on the chromatin landscape and the looping of the DNA may have differential expression that has not been accounted for.

Lastly, identification of biologically significant regions using only transgene analysis requires complementary research using knock-out studies as transgenic studies. This is because the enhancer screens using transgenic mice that we have used in this study do not tell us if these regions are functionally relevant or if they play an important role in the regulation of the gene we believe they are targeting. One major consideration with the use of mice is they may not mimic human physiology exactly but does give us a good indication of what is happening.

5.2. In silico annotation of genes for enhancers requires functional validation

With advancement in technology and computing power there has been a plethora of data generated for gene regulation in terms of ChIP-seq targeting TFs and histone

modifications, DNase1 hypersensitivity, micrococcal nuclease digestion followed by sequencing (MNase-seq), Formaldehyde-assisted isolation of regulatory elements followed by sequencing (FAIRE-seq), Assay for transposase-accessible chromatin using sequencing (ATAC-seq), 3C, 4C, 5C, Hi-C, ChIA-PET and (Meyer and Liu, 2014, Belton et al., 2012). These techniques provide information on the chromatin state, the occupancy and the conformation of the DNA. With ENCODE the data for each of these techniques are publically available to researchers to analyse and identify region of interest to examine functionally. Histone marks have been regarded an indicator for enhancers and other regulatory elements in the genome, with active enhancers highly enriched in H3K27ac. Histone modifications H3K27ac, H3K4me1 and H3K4me3 along with evolutionary conservation was used in this study and found that these were good criteria for candidate enhancers within the *Acan* locus. However, not all enhancers were highly conserved, the -65kb region could not be found in the opossum or fugu sequence whereas the other (+60, +55, +26, -10kb, -35kb and -87kb) are conserved to a minimum of opossum. There were further sequences within the locus that were not examined but held good indicators for enhancer activity that cannot be ruled out to having a function. High conservation of DNA and RNA have shown tissue specific similarities between close and distantly related vertebrate species and tissue specific expression patterns can be seen despite sequence divergence to larger sequences (Pennacchio et al., 2006). However, conservation in non-coding regions of intergenic and intronic regions of genes do not always correlate with gene specific regulatory elements, although a good indicator to functionally relevant regions, as ChIP-seq experiments are able to detect functionally active *cis* acting elements that are not evolutionary conserved and do not confer biological advantages (Birney et al., 2007).

The use of LOD scoring and the table browser utility of the UCSC genome browser only gave highly significant regions that were conserved and held histone modifications resembling that of enhancers. This may be problematic technique to use, as the enhancer definition evolves with new and more advanced techniques, for example H3K4me3 generally mark promoters, but the -65kb held this enrichment with no

mapped gene TSS within its vicinity. It has been shown, at least in cancer cells, that the H3K4me3 modification can mark enhancer activity, therefore this mark should not be excluded from enhancer screens (Proudhon et al., 2016). LOD scoring is based on the prediction that is user defined and could potentially exclude regions of genuine interest.

It is interesting to note that when we performed liftover from the mouse to human and mapped the enhancers that in the 9 cell lines used in the chromatin state studies that only two *Acan* enhancers gives signals of enhancer activity, the +26kb is marked as a weak enhancer in HMEC cells but an insulator in the other cell types. Additionally, the -10kb in H1-hESC, HSMM and NHEK is marked as weak enhancers but has repressive marks in the other cells types. The -35kb is marked as a promoter, the -65kb is weakly transcribed and the +60kb, +55kb and -87kb are repressed (Figure 19). These predications give an indication of what the chromatin state is like in other cells and allows a picture to develop of the function of these enhancers in non-chondrogenic cells, they do not however, provide any indication of the function of these regions.

For simplicity the approximate positions of the +60, +55, +26, -10, -35, -65 and -87kb were taken as the start of the region examined to the TSS, even in shorter constructs the name was kept to avoid confusion.

5.3. The multiple chondrocyte expressing enhancers of Acan

Cis-acting elements such as enhancers have the ability to control temporal and spatial expression of gene in a tissue specific manner irrespective of their position and orientation, the mammalian genome appears to have an abundance of regulatory elements compared to coding sequences (Bulger and Groudine, 2010, Shen et al., 2012). It is becoming more apparent that there are more genes under the influence of multiple enhancers that dictate robust expression and responsiveness to cellular and physiological stress, particularity the imprinted genes and the Hox gene cluster that allow co-ordinated limb patterning (Verfaillie et al., 2016, Riso et al., 2016, Gonzalez et al., 2007, Tschopp et al., 2011, Spitz et al., 2003).

ACAN is an important molecule for the normal physiology of the developing and adult cartilage. Understanding the transcriptional inputs that direct *Acan* expression in the chondrocyte will provide critical information that brings about its expression under normal condition and provides avenues to explore its dysregulation and loss in diseases such as OA and genetic linked chondrodysplasias. Previous studies have demonstrated the role of Sox9 along with Sox6 and L-Sox5 that directs the expression of the -10kb enhancer (Han and Lefebvre, 2008). However, this enhancer appeared incomplete hinting of the existence of other regulatory elements. Zebrafish studies have offered a more comprehensive insight into the *Acan* enhancers with other studies suggesting a high level regulator upstream of the known enhancer (Hu et al., 2012, Ikeda et al., 2014). We find two enhancers capable of driving tissue-specific expression of the chondrocytes found upstream of the TSS. One within the first intron of *Acan* and one at -35kb. These enhancers drive a more global chondrocyte phenotype compared to the known, -10kb, which is confined to the core regions. It is curious to note that we used limb specific criteria for selection of enhancer elements that were not chondrogenic but rarely did we see expression outside of the chondrocyte. The reason for the specificity of the enhancers may be detailed by the TFs that dictate the cell-type expression or the chromatin environment in the chondrocytes compared to other cells types.

The intronic sequence (+26kb) and the -35kb up stream element are able to recapitulate *Acan* expression at E15.5 occurring in all chondrocytes irrespective of their maturation stage suggesting they might direct *Acan* transcription on a basal level, a summery is described in Figure 76. It is known that enhancers rarely act alone; they can have a shadow or other enhancer that co-operate to bring about expression (Werner et al., 2007, Wenger et al., 2013). The onset of *Acan* expression is around E10.5 in the mouse, a time point that was not examined in this study, it may be of interest to examine these two enhancers at earlier time points to determine if they are supporting each other, having additive effects, or responsible for timely activation of *Acan.* The -35kb, +26kb and the -10kb human sequences were identified to express in the zebrafish enhancer detection experiment showing that their conserved sequences can drive similar

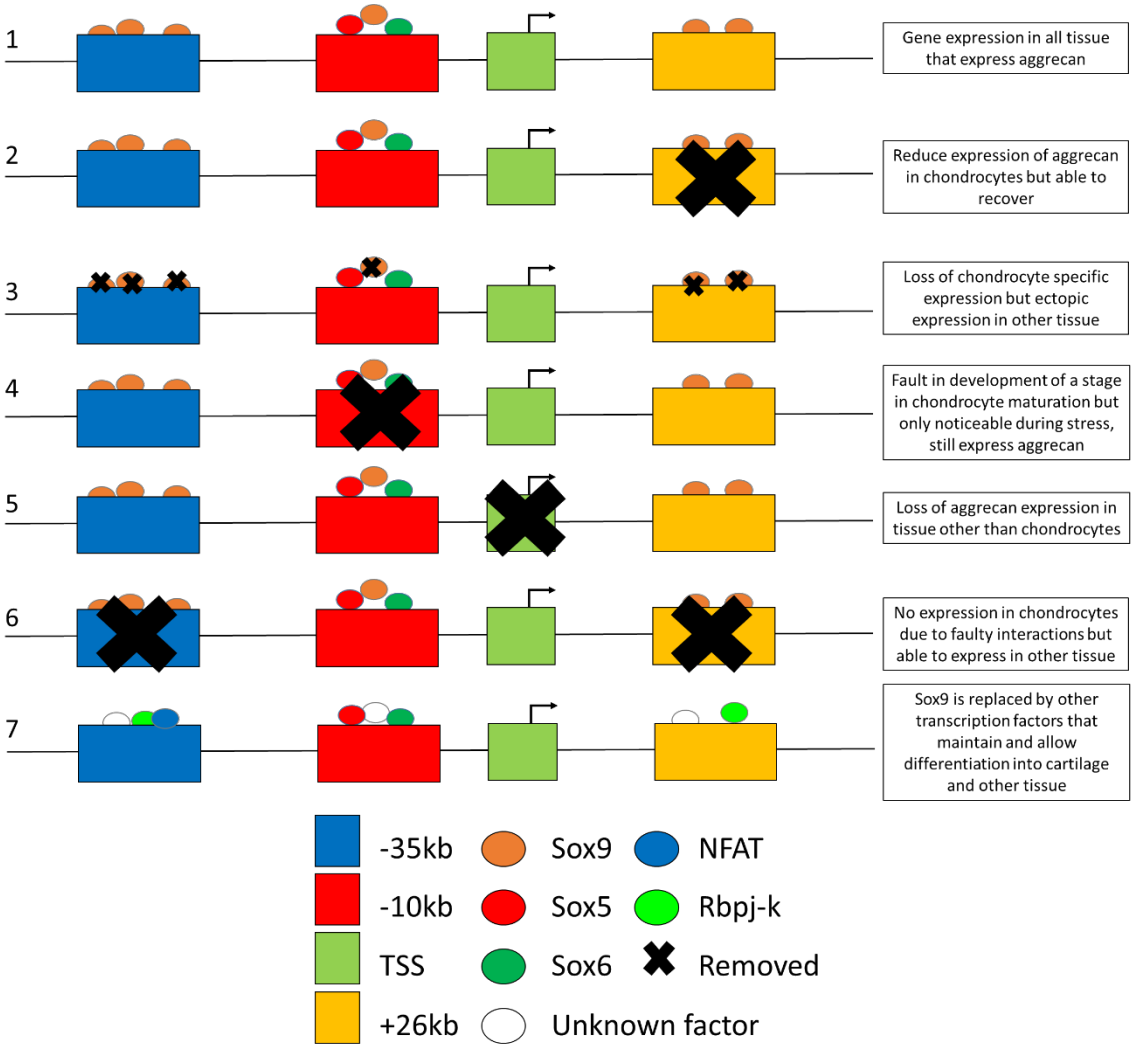
expression in other animal models (Hu et al., 2012). The consistent overlap in expression between the +26kb and -35kb suggests that there is redundancy or that each enhancer allows the activation of *Acan* at distinct stages of development with a knockout of one enhancer delaying the onset of expression, this hypothesis would need to be examined and can be done so using CRISPR/Cas9 technology (Huang et al., 2016). Other enhancers identified in the screen such as the -108 and -115kb may contribute to chromatin landscape and have additives effects on the other enhancers, but in our screen inactive enhancers and enhancers from other cell types were not considered. Other cell type specific enhancers may add to the reservoir of regulatory elements that are clustered in close proximity of *Acan*, as we have not found regions that drives expression in the brain and heart valve. It may be that the enhancers they are found to regulate the chondrocyte expression are transiently active in busts or short periods of time in other regions that are missed by our reporter method.

Furthermore, inactive enhancers in one cell type may be active in another, these serve to allow the effects of the other enhancers in a lineage, stage specific manner as well as allowing the local and distal effects to be corresponded into transcriptional output. For example, in B lymphocytes the *IgK* enhancers and the proximal enhancers 3'Ek, MiEk, Edk and Ck and Jk are active but in T lymphocytes they are inactive, using 4C-seq baits, show that each enhancer contributes to creating an transcription hub and rearrangement of the chromatin, but a removal off one leads to dissolution and reduction of transcription, not a complete loss (Proudhon et al., 2016). This study identified the inactive MiEk enhancer cluster significantly influence T cell receptor beta locus recombination (Proudhon et al., 2016). Therefore, the repertoire of enhancer function is ever growing and the *Acan* enhancers identified in this study may have unpredicted functionality.

ACAN is stably expressed and turned over in adult cartilage and enhancers are generally explored in diseases such as cancers and in development with adult normal physiology being neglected (Chambers et al., 2002). We wanted to examine if any of the *Acan* enhancers are expressing in adult mice, however, we never interrogated its function in

the adult tissue. The -35kb enhancer replicate ACAN expression in articular cartilage in 8 week and 6 month old mice. The relevance of an enhancer at this stage may allow binding of regulatory TFs to allow turn over, however, without further experimentation the role of the enhancer in adult cartilage cannot be certain (Kohn et al., 2012, Mirando et al., 2013).

Figure 76: Schematic of the role of the -35kb and +26 *Acan* enhancers



A hypothesis that arises from this body of work is that the -35kb +26kb are redundant, together along with the promoter and the -10kb there is normal expression of Acan.

When a mutation occurs in either the -35kb or +26kb there is reduced expression but the system is able to recover therefore not detrimental to transcription, but the loss of Sox9 removes the chondrocyte expression of Acan driven by the enhancers. In the absence of the -10kb the other two enhancers drive normal expression of Acan but the -10kb is there to protect against stress or for timely expression of Acan during certain times of development. In the absence of the promoter there is a loss of expression of Acan in certain cells types but as long as the +26kb and -35kb are intact chondrocyte specific expression is not hindered but if both of these enhancers are lost there is defective chondrocyte expression. Finally, in adulthood Sox9 is replaced by other transcription factors such as NFAT, Rbpj-k or other TFs to maintain Acan turn over.

5.4. The role of the -65kb and -87kb enhancers

As mentioned before, it is known that enhancers rarely act alone; they can have a shadow or other enhancer that co-operate to bring about expression (Werner et al., 2007, Wenger et al., 2013). Of the two far up-stream enhancers, the -87kb appears to be temporally expressed, with our earliest detection in the chondrocytes was at E15.5. At this age the signal is weak, but appears to be following the interzonal regions of the joint, as the mouse ages the staining remains primarily in the articular cartilage but not the IVD, analysis of this enhancer during joint formation may shed a light on its functional role. It is known that at E14 mRNA of *Acan* is detected in the mouse's tibial cartilage in the metaphysis and the epiphysis but weakly in the hypertrophic chondrocytes of the diaphysis (Shibata et al., 2003). Therefore, to find a regulatory element that appears solely in the hypertrophic chondrocyte such as the -65kb was unusual but not unexpected, as the *Acan* knock-in mouse also showed hypertrophic chondrocyte expression when the cre was activated (Henry et al., 2009). The precise-timing of the morphological event in long bone formation and growth is important with the loss of Acan results in a smaller and disorganised hypertrophic zone in the growth plate of NMs, which suggest that an enhancer in the hypertrophic chondrocytes may have some importance and serves a different role as opposed to regulating

transcription, but rather the organisation of the chromatin. What we must also consider is the half-life of the β-gal which can range from minutes to greater than 20 hours *in vivo* depending the pH and temperature (Bachmair et al., 1986, Rojas et al., 2005). It maybe that the -65kb was active at an earlier stage of development but is inactive at E15.5 but the enzyme remained in the tissue, therefore the staining is an indicator of the cells progressing through the cartilage (Domowicz et al., 2009).

In the zebrafish study the -65kb falls outside of the examined enhancers and the -87kb failed to express (Hu et al., 2012). The -65kb in fact when examined closely on the predication is not highly conserved so may not have the relevant machinery or available TFs in the zebrafish to bring about its expression. The -65kb is of interest to us mainly because of its different histone profile and DNase1 hypersensitivity compared to the other enhances identified in this study, the presence of the H3K4me3 and the fibroblast DNase1 peak indicates that the enhancer is highly active, the explanation of which may be more complicated than TF interactions but due to bending of the DNA and the promoter-enhancer interactions that have not been explored.

The -87kb expression in 8 week old mice is impressive, staining in the growth plate and the superficial lay of the articular cartilage are a lot stronger when compared to the expression at E12.5 and weak expression at E15.5. This shows that this enhancer is (at least in part) responsible for adult expression of ACAN and may have a key role in the adult transcriptome of this gene. With this idea that the -87kb is adult cartilage specific it may mean that cells that are marked by the -87kb enhancer in early development are a different chondrocyte to other chondrocytes marked by the other *Acan* enhancers and have already committed to a lineage fate. Therefore, the -87kb may drive the lineage specificity towards the growth plate chondrocyte and into the superficial zone. Compared to the -35kb enhancer, the expression is almost identical, albeit the -87kb being strong at this stage. This may be due to the copy number and insertion site. The expression of both the -35kb and -87kb were in only the superficial zone and rarely in the deeper regions although ACAN is found in the deeper zones. This raises questions, is

the penetration of the X-gal poor or alternatively, is there a TF that drives expression in chondrocytes and restricts the enhancers to the superficial layers. Research that supports this idea of a TF is the work in lubricin (*Prg4*) which is expressed primary in the superficial zone chondrocytes of the articular cartilage (Reesink et al., 2016). The *Prg4*-CreER^{T2} is an inducible cre knocked into exon 2 of *Prg4*. With immunofluorescence detection of the cre and X-gal staining of *Prg4*CreET^{T2} crossed with *Rosa26*^{floxLacZ} mice show that *Prg4* is expressed in the superficial a region at E17.7 and during development it remains in this region. However, in adult mice when lineage tracing was conducted for 30 days, 6 months or 18 months show that these chondrocytes give rise to approximately 70% of the superficial chondrocytes but there are cells migrating into the deeper zones of the cartilage as opposed to remaining in the superficial region. The study suggests that these chondrocytes are progenitors to deep zone chondrocytes contrary to the dogma of chondrocytes expanding towards the superficial region of the cartilage (Kozhemyakina et al., 2015a, Kozhemyakina et al., 2015b). This hints that there is a restriction to the superficial zone placed on these chondrocytes and there is a trigger that allows them to migrate down. One of these triggers could be TFs that help retain the expression of the gene to the superficial region. Therefore, one can infer from this study that the two *Acan* enhancers (-35 and -87kb) may be under the same influence of this unknown TF. What is needed is to establish what is the regulation that retains the -35kb and -87kb to the articular cartilage and to the growth plate and if it is the same in the different regions. To do this one suggestion, albeit costly, is to fragment the enhancers and determine the minimal region that regulates expression and perform a TF platform assay. Alternatively, ChIP-seq or DNase1 hypersensitive can reveal sequences of TF binding that can be aligned to the TF databases to identify possible factors to be examined.

5.5. 3' intronic sequences found within the Acan locus

When the intronic regions of *Acan* were examined it was noted that closer to the 3' region of the gene there were less limb specific histone modifications. This may be

because the screen for histone modifications in the intergenic regions were over looked. The +60kb only had H3K4me1 suggesting it was poised whereas only weak/ poor peaks for the histones were seen at this stage of development where the histone ChIP-seq were performed. However, its expression at E15.5 was tissue specific in individual embryos but different when compared to each other. These regions rarely share similarities, although two do share some chondrocyte expression, this region was highly conserved down to fugu suggesting that evolutionary this region has withstood genetic stresses, hinting that this region has a beneficial purpose to the host, what this is remains to be examined. . Looking at the sequence, it is small at 411bp, but could well house multiple TFs binding sites for different tissue types or the sequence is missing essential regions that may lie in the exons 16 and 17. However, its role in regulating *Acan* will require further investigation.

The +55kb is a non-chondrogenic enhancer with lung, kidney and tissue specificity with expression at E15.5, 8 week and 6 month old mice showing nucleus pulposus staining. The specificity of the kidney staining in the glomeruli makes it unlikely to be due to endogenous *LacZ* staining, however *Acan* has not (as yet) been shown to express in the glomeruli and the function in this area is unknown (Sebinger et al., 2013, Weiss et al., 1999). Only other members of the Lectican family such as VCAN and BCAN, appear to be critical in the kidney (Rudnicki et al., 2012, Naso et al., 1995). In terms of the lung ACAN was not found to express, rather its family member VCAN has been localised to this area. Areas where ACAN had been detected outside of the cartilage include the aorta, brain and skin and it is interesting to note that none of the enhancers identified in this study had been found in the aorta which may have been over looked (David et al., 1989).

The tissues containing +55kb expression all derive from different lineages, the lungs from E9 arise from the ventral foregut endoderm, glomeruli from the mesoderm (Abrahamson, 2009) and the nucleus pulposus is derived from the cartilage end plate into fibrocartilage (Kim et al., 2003, Jeong and Epstein, 2003). The different tissues leads

to the question of what links these three tissues if one enhancer can express in these regions and what drives its expression? This may be due to the size of the +55kb (around 2.5kb) which could be split into three regions that can drive expression only in the kidney, lung and nucleus pulposus?

To examine the last question, ENCODE data for histone modification in these regions can be obtained and examined. There are no signal for this area for any histone modifications in the lungs, but in the kidney there is a small peak of around 512bp, this is most promising, although it supports the idea that this enhancer may regulate these areas simultaneously (Figure 77). This leaves this gene specificity and minimal sequence of activity of the enhancer open for debate and requires further dissection to understand its true role in transcriptional regulation.

Figure 77: Analysis of the +55kb in other tissues

The large +55kb region examined for histone modifications H3K4me1, H3K4me3 and H4K27ac in the lung and kidney with the limb and fibroblast overlaid. The +55kb is generally void of histone modifications but in the kidney highlighted in red.

5.6. Regulation of Acan enhancer activity by Transcription factors

Enhancers serve as TF binding sites for co-ordinated transcription activation and regulation, Sox9 is known as the master regulator of chondrocyte fate so it was unsurprising that binding motifs were found within these *Acan* enhancers. We find that Sox9 binds to all the chondrocyte expressing enhancers although not always a class II binding. During the analysis of these regions publications of Sox9 ChIP-seq data became available (Ohba et al., 2015, Liu and Lefebvre, 2015, Garside et al., 2015). We used these to validate our EMSAs, by lifting over and converting the data into the mouse we layered each dataset to visualise where Sox9 binds. The *Acan* locus appears to interact with Sox9 frequently in both rat and mouse studies. This is true for our enhancers, the +26kb, -35kb, -65kb and -87kb all show Sox9 binding from the rat and class I and class II, this is different to the -10kb for which only the rat data supports Sox9 binding. The +60 and +55kb enhancers do not have regions of Sox9 binding which strengthen the non-chondrogenic potential, but raises doubts for the chondrocyte expression of the +60kb and if it is under the influence of another TF. Furthermore, the enhancers identified in our study are the only regions to interact with Sox9 in mouse limbs at E12, this may suggests differences in the threshold set by the ChIP-seq studies or that the interactions in cell culture are different from in whole limbs (Garside et al., 2015).

Similar to the known enhancer these regulatory elements interact with Sox9 although not all of the enhancers engage in the suggested Class II binding of the Sox-dimeric motif. It may be that for -65kb and -87kb they interact through the less preferential Sox monomer and require an integrated input of other transcription factors (Ohba et al., 2015). This leaves the question of why there are differences in the expression of the *Acan* enhancers when Sox9 appears to regulate them all, at least *in vitro*. To find Sox9 binding interactions with the -65kb enhancer is intriguing, as over expression of Sox9 leads to a reduction in chondrocyte maturation into hyper-trophy and a loss of Sox9 accelerates this process, suggesting Sox9 has a repressive role in hyper-trophic chondrocytes (Akiyama, 2008, Akiyama et al., 2002, Akiyama et al., 2004, Bi et al., 2001). This leads to the idea that Sox9 binding to the -65kb may confer the enhancers

specificity to the chondrocyte but requires the action of other TFs, that may act synergistically or competitively with Sox9 to drive the hypertrophic part of its repertoire. Other studies support this notion with Sox9 being shown to have a dual role in repressing and activating transcription, it represses *Col10a1* by direct interaction in growth plate chondrocyte to limit *Col10a1* to hypertrophic cells only (Leung et al., 2011). In a different case, Sox9 directly binds to *vegfa* in cultured primary chondrocytes to subdue the expression and inhibit vasculature formation and suppression of terminal differentiation (Hattori et al., 2010). Additionally, work screening 15kb up stream and 10kb downstream of the CTGF gene (*CCN2*) in the RCS cell line using ChIP-on-chip for Sox9 found that the binding site was not the consensus binding site, but five CATTcAg and not as the dimeric motif at the proximal promoter (Oh et al., 2016). This could underline differences in regulation of the enhancers and the role of Sox9 and its differences in binding preferences. What this may mean is the binding preferences of Sox9 to the different enhancers may confer different activities, the stronger class II binding may be strong enhancers and the weak or more dynamically controlled enhancers require other TFs to function. This of course requires further examination to determine if the hypothesis is true.

We wanted to examine the importance of Sox9 in regulating the *Acan* enhancers. To do so we took the -35kb element and subjected it to mutagenesis studies. Sox9 binding site mutations did not attenuate the transgene expression *in vivo* rather the expression was abolished from the chondrocyte and expression was generated in the periosteum or fibroblastic cells. This result supports the motion that Sox9 is required for the chondrocyte specificity and represses the expression in non-chondrogenic cell lineages. The shift indicates that other transcription factors are able to regulate the expression of the -35kb. The perichondrium expression has previously been noted in the Acan mutated chick(NM). In this model a loss of aggrecan core protein resulted in a shift in regulatory gene products such as growth factors and signalling molecules particularly in the hedgehog effector Patched (PTCH) from the developing periosteum suggesting that a fault in *Acan* was able to reproduce a perichondrium effect (Domowicz et al., 2009).

The mechanism that is involved remains unexplored but an evaluation of the knock-out of the Sox9 binding sites *in vivo*, by the use of genome editing techniques, may help uncover what is happening to the expression of the -35kb enhancer Sox9 mutated sequence and in the NM chick. When mutations are introduced in the three dimeric Sox9 binding sites independently there is a loss of chondrocyte specificity with each one. This questions if these three sites are interacting or stabilising the Sox9 binding to the enhancer. Additionally, none of the mutations individually reveal periosteum expression and site 7 retained some chondrocyte expression, which may suggests its redundant role in comparison to the other two binding sites, or that another transcription factor has activated that subset of cells. The single site mutations suggests that the removal of Sox9 binding allows or prevents a different protein from binding, as mutations in each site still allows a minimum of two other Sox9 binding sites, this may mean that there is occupancy around the regions blockading other TFs but not significant enough to drive the expression in the chondrocytes or that the DNA is looping in a direction to reduce binding of other TFs this idea requires further verification. Site 3 presented with some fibroblast expression in one transgenic embryo, which we cannot rule out a SOI effect. What is certain all 3 binding sites intact are required for the normal function of the -35kb enhancer? What sox9 is doing at each site is to allow the remodelling of the chromatin to initiate transcription with the loss of all three sites the remodelling cannot occur so the chondrocyte specific expression is loss (Coustry et al., 2010).

One possible TF that could be interacting is the Notch transcriptional effector Rbpj-κ, to determine if rbpj-κ is able to regulate the +26kb and -35kb as it does with Sox9 (Kohn et al., 2015, Mead and Yutzey, 2009), three possible rbpj-κ binding sites was identified in the -35kb sequence, with one binding *in vitro* and three binding sites in the +26kb, with sites 1 and 3 binding. Mutations *in vitro* we are able to abolish the binding in both enhancers binding sites apart from site 1 for the +26kb, which appeared to still bind albeit to a lesser extent. This may be due to the binding motif for Rbpj-k although not as strongly predicted is still partially intact in the mutant oligo at the site: *CTTCCCAGGG* when compared to the consensus (Figure 28 C). When the mutation to site C of the -

35kb was introduced *in vivo* we saw very little change in chondrocyte specific expression of the transgene. In some cases the expression in the chondrocyte was scattered but still remained strong, in many of the transgenic embryos there was an increase in the expression in the dermal fibroblasts and there were expression in the blood vessels. It may be that the mutation in the Rbpj-κ site has changed the binding dynamic of other TFs at the -35kb enhancer but the change isn't significant to result in a complete change in expression of the transgene. Rbpj-κ is required for terminal differentiation and regulates endochondral ossification and is necessary for the maintenance of joint and articular cartilage with a knockout of *Rbpj-κ* resulted in a reduction of *Acan* mRNA. Therefore we may not see an effect of our mutation until later stages of development and in adult mice. Alternatively it may only occur with both Rbpj-κ and Sox9 binding site mutations (Kohn et al., 2012, Chen et al., 2013, Kohn et al., 2015, Mirando et al., 2013). The similarities between the Sox9 and the Rbpj-κ mutants are the dermal fibroblast expression. Rbpj-κ mutations may expose a binding site for dermal fibroblast expression of the enhancer, or it could be an inhibitor of the expression of the transgene in other cell types. A diagram of the role of Sox9 and other TFs in different cell types generated from our observations and the literature is detailed in Figure 78 and Figure 79 to try and understand the role of the TFs in development and in the adult. With transgenics the transgene integration is random, the copy number is variable in each case and with the transgene theoretically being present in every cell type, there can be expression of transgene in regions that is not predicted. In the WT sequence of the -35kb there is always robust chondrocyte expression with little or no expression in any other cell type. The regulation by the transcription factors and various signalling pathways such as sonic hedgehog (ssh) allows for Sox9 and another transcription factor Nkx3.2, which represses the Gata proteins (Daoud et al., 2014), to establish the chondrocyte phenotype from the sclerotome, mesoderm and neural crest (Zeng et al., 2002, Liao et al., 2014, Akiyama et al., 2005, Akiyama et al., 2004). In other cell lineages there are transcriptional activators and repressors that dampen or intensify expression and activity of Sox9 to allow or prevent the differentiation into other cell types such as tenocytes (Sugimoto et al.,

2013a), osteoblasts (Zhou et al., 2006, Dy et al., 2012, Goodnough et al., 2012) and vascular smooth muscle cells (vSMC)(Briot et al., 2014). During cranial development while differentiation of the dermal fibroblast occurs, Wnt/ β-catenin influences dermal cell differentiation by regulating the TF *TWIST2* via Lef1 and TCF4 (Goodnough et al., 2014). TWIST2 is a transcriptional repressor that has been shown to suppress *Sox9* mRNA (Tran et al., 2010). This may support the idea that the enhancers serve as one of the platforms for the signalling and TFs that down regulate the expression of *Acan* facilitating the differentiation of the stromal cells into dermal fibroblasts rather than chondrocytes. An approach to identify other biological important TFs that would eliminate the requirement fors identifying the sequence of the TF before functionally testing, is to carry out panel screen of enhancers relevant to the cell of interest. This will identify some of the TFs that bind and the possible combinations of TFs that bind in close proximity for further analysis, and the sequence can be pulled out using native ChIP. This, however, does not allow the identification of novel binding factors (Huang et al., 2016).

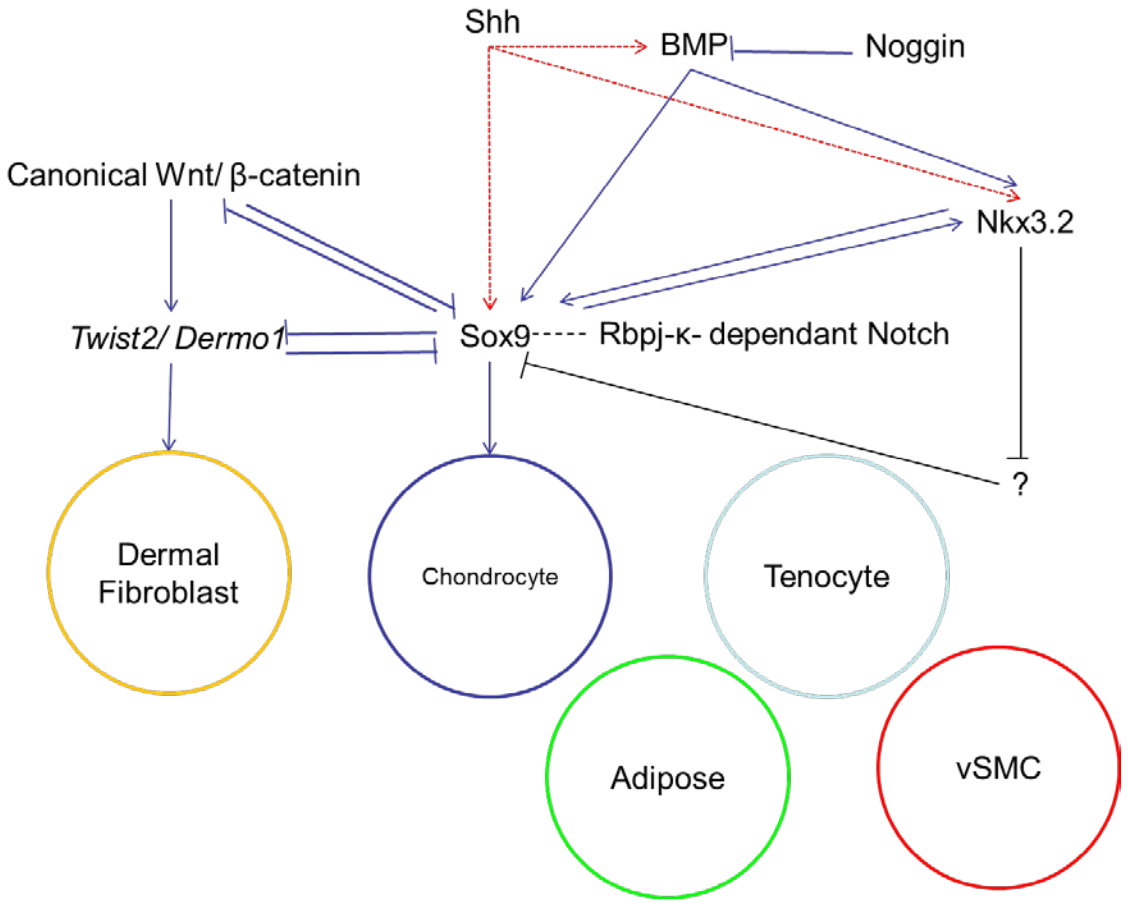
We have yet to establish other TFs that are interacting with these *Acan* enhancers and if initiators or more core PIC complexes form at these enhancers, but the literature does suggest TFs AP-1, RUNX2 (Zhou et al., 2006), Mef2 (Gu et al., 2014, Arnold et al., 2007) family and Hox, which may aid Sox9 in enhancer activation and we have not explored the role of L-Sox5, Sox6 (Dy et al., 2010) or even the SoxC proteins (Dy et al., 2008) that have been shown to regulate different time points in chondrocyte maturation. NFAT1 was predicted to bind to the +26, -35 and -87kb enhancers and we have shown that the protein binds *in vitro* to these constructs. For *in vivo* analysis were generated, but due to time constraints the experiment was not conducted. It is interesting to note that the enhancers share the same TF binding, although the function of each one has not been explored. A question which arises from this is that if all these TFs bind to the same enhancers and drive the same specificity what differentiates the chondrocytes from other cell types? There is a need to explain on a basic level, why TFs tissue specific knockout are so detrimental to the entire transcription and translation machineries

which leads to poor expression in only that cell type, not just in every cell that both the enhancer and TF are found. A question to ask is if Sox9 is present at all enhancers what are the factors that are interacting with Sox9 to create the differences seen in these *Acan* enhancer? And if these form stable transcription hubs to allow transcription only in the chondrocyte?

A working model of the role of TFs competition at enhancers in different cell types in both development and in adult maintenance is depicted in Figure 78 and Figure 79.

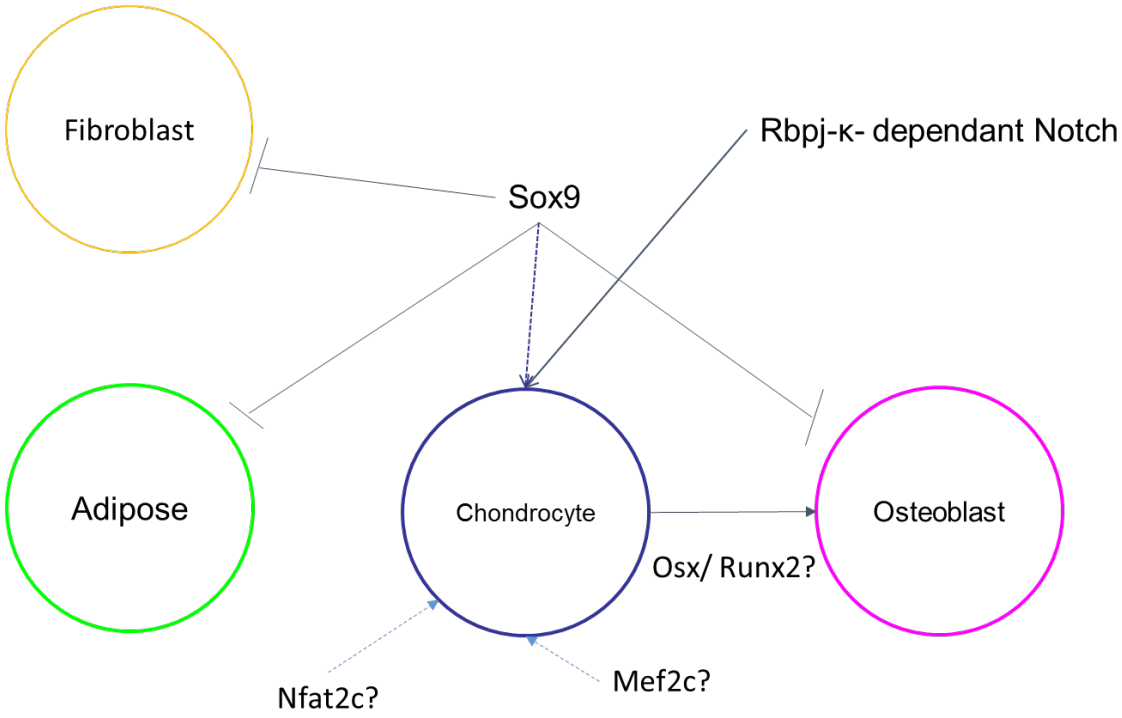
.

Figure 78: Predicted diagram of the role of Sox9 in regulating enhancer expression



Assumptive diagram of lineage expression of the enhancers control by sox9. Signalling pathways regulate Sox9 such as Shh and BMP, other transcription factors such as nkx3.2 and Twist two compete may away Sox9 binding in the area around enhancers to drive the precursors to other cell types such as the fibroblast, adipose, vascular smooth muscle or tenocytes. One hypothesis is when Sox9 is expressed in more abundance it binds to the enhancers that are chondrocyte specific and allows the chondrocyte expression of the gene regulated by the enhancers that serves as platforms for further engagement of extracellular matrix and chondrocyte specific genes.

Figure 79: Predicted role of Sox9 in adulthood



It may be in the adult system Sox9 acts as a regulatory to keep enhancers expressing in the chondrocytes but other transcription acts regulate their role and the effects they have on turn over or activation of certain Acan enhancers. It is possible that the

introductions of other transcription factors or the removal of sox9 leads the expression to divert into other cells.

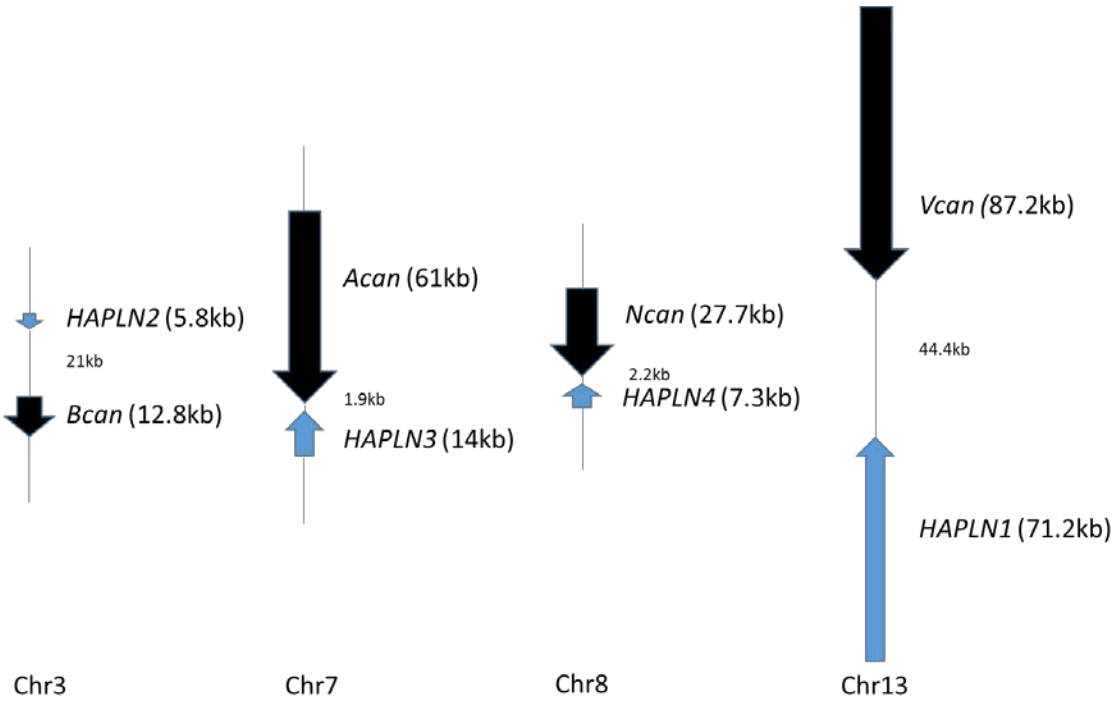
5.7. Evolutionary co-regulation of link proteins and Lectican protein family

The proteoglycan family lecticans and the link proteins appear to have evolved evolutionary together, splitting from a core transcript early in evolution like the *HoxA-D* genes (Figure 80) (Spicer et al., 2003). It has been proposed that due to the close proximity on a genomic scale to the proteoglycan transcript to the link proteins that they are co-expressed and act on each other during the formation of the proteoglycan aggregates. It is known that HAPLN1 interacts with both VCAN and ACAN although through different domains, this may mean that the other HAPLNs may aid in the formation of these aggregates in other tissues (Matsumoto et al., 2003, Shi et al., 2004). Currently, there are very few experiments that link enhancers to the target gene generally they are linked to the nearest gene, *HAPLN3* is genomically linked to *Acan* and is the most widely expressed link protein, although it is not expressed in chondrocytes. The knockout mouse have not been generated, from our data an interesting proposition and observation can be put forward. Due to the proximity of the genes and that they have emerged together in evolution, transcriptional regulation of the smaller gene, *HAPLN*, may arise from the larger proteoglycan. When the division of the major gene the transcriptions may have come from the same promoter and evolutiongave rise to the 4 different links and proteoglycans to allocate their tissue specificity, with Bcan and Ncan being neuronal specific but at different times at evolution (Figure 81). Genetic events results in a greater distance between the genes to allow for robust regulatory elements in different tissue types. Therefore, it may not be so farfetched that enhancers that lie within one gene regulates another gene to direct the tissue specific expression or that like bi-directional promoters that drive activity of neighbouring genes to allow coordinated expression of both genes, this may also occur in enhancers (Ong et al., 2002, Trinklein et al., 2004). This idea can be examined using bi-directional luciferase vectors and using 3C, 4C, 5C, Hi-C and *in situ* Hi-C technology to identify the gene in

which the enhancer identifies with (Jin et al., 2013). The reason for suggesting this possibility is that the intronic enhancers identified in this study, the +55kb, expresses in an area of the kidney not previously shown to be linked with ACAN but a where *HAPLN3* is detected by northern blot analysis (Spicer et al., 2003). It must be noted that enhancers can act on different genes in different cell types, for example the HS40 enhancers of the β-globin gene also acts on the epsilon-globin gene (Bender et al., 2006, Bender et al., 2012, Chen et al., 1997, Nickol and Felsenfeld, 1988). Additionally, the +60kb region drives expression differently in different transient expression at E15.5, which may be down to SOI effects and the sequence requires shorting or protecting, but these areas are where *HAPNL3* is expressed and the enhancer acts as a signalling/ TF binding platform along with other enhancers to drive both *Acan* and *HAPLN3* expression. These are all assumptions but are interesting discussion points. The intronic enhancer is unlikely to regular *Acan* as IHC does not localise ACAN to areas where the +55kb enhancer is expressing giving an indication that it is more likely to regulate solely HAPLN3, indeed HAPLN3 expresses both in the kidney (both tubules and glomeruli) and in the lungs when the tissues is examined in the human protein atlas (http://www.proteinatlas.org/ENSG00000140511-HAPLN3/tissue). If the intronic enhancers regulates *HAPLN3* rather than *Acan* then the regulation of the chondrocyte specificity of *Acan* is determine by the intergenic enhancers and the 5' intronic regions such as the -35kb enhancer, (Figure 82), this can be examined using knock-outs of the enhancers. This could mean that the +55kb allows *HAPLN3* to be activated at the same time as *Acan* transcription but in different tissue or the enhancer may not need *Acan* transcription to occur to be active. The hypothesis is greatly weakened when, data from unpublished gene array provided by Professor David Young and Kat Cheung from the University of Newcastle, show when MSCs undergo differentiation into chondrocytes *Acan* is transcribed but *HAPNL3* is not, looking at the other lectican protein family *Vcan* is switched off when the cells become chondrocytes whereas the accompanying link protein *HAPLN1* increases its expression putting doubts at least in chondrocytes the co-regulation of the links and lecticans (Figure 92). However, in the whole mount the

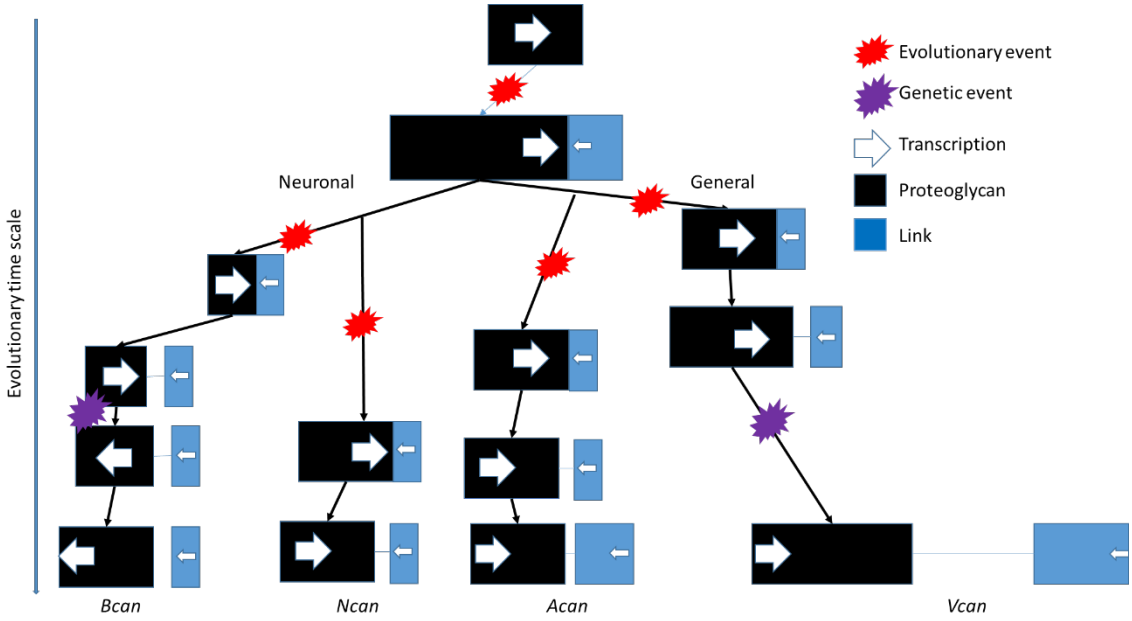
tissues are different in expression, Acan is mainly found in the chondrocytes and rarely in the kidneys, our BR1 staining even shows that it is the tubules that express Acan. Does this mean that the +55kb regulates the HAPLN3 in the glomeruli irrespective of the enhancers host gene? Additionally, our antibody staining using HALPN3 antibody presents weaker expression in the mouse compared to the human data, it only expresses in the medulla and in the lungs only in the bronchioles, and this may be due to the antibody being raised for the human antigen. Our antigen retrieval using hyaluronidase rather than citric acid as they do in the human protein atlas may account for the differences, this leaves a very big questions of what the +55kb enhancer is regulating, we cannot rule out the possibility that the enhancer of regulates either *Acan* or *HAPLN3* since we only utilised two antibodies, the transcript may be an isoform or short lived. Proteomics or knock-out experiments are required to explore this element in detail. The *Bcan* and *HAPLN2* promoter direction may have occurred due to genomic rearrangement but overall the *HAPLN* and lectican co-regulation offers an interesting avenue for how enhancers and other regulatory evolved and if co-regulation occurs. Additionally, examining the role of the *Acan* enhancers on expression of the *HAPLN3* may yield a functional role for each regulatory element, it may not be inconceivable that there may be cross-talk between the regulatory elements of two genes that share similar genomic space..

Figure 80: The distance between link proteins and the lectican family



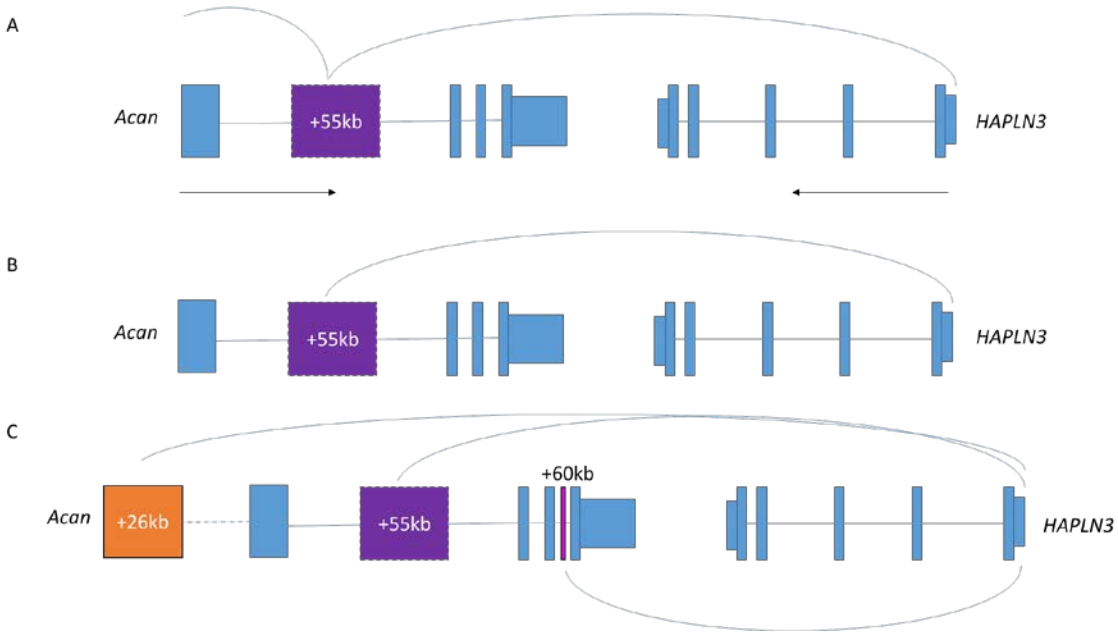
Size and distance apart of the Lectican family members; Bcan, brevican; Acan, aggrecan; Ncan, neurocan; Vcan, versican, HAPLN1-4 are the link proteins 1-4.

Figure 81: Evolution of the link and lectican proteins



Schematic of e lectican family and link protein evolution. It may be that in evolutionary time that these members evolved at different times, with Vcan being the oldest due to its widespread expression, Acan being next and Bcan being the youngest. This may hint that their regulation evolved together through both evolutionary and genetic stresses.

Figure 82: Possible interactions between enhancers found in the introns of *Acan*



Schematic of the 3' Acan enhancers and possible reason for their interaction that lead to their expression. The +55kb could be bi-directional and regulate the HAPLN3 gene as well as Acan due to the proximity of the enhancer to both genes (A) or the +55kb could have formed to regulate HAPLN3 so the expression of both genes are switched on at the same time, not necessarily in the same tissue . +55kb may not be the only enhancer to act on HAPLN3 as the two other intronic enhancer of Acan (+60kb and +26kb) are within the TAD of HAPLN3 and could play a role in its regulation.

5.8. What are the functions of multiple Acan enhancers?

The upstream elements,-35kb, -65kb, -87kb and the known -10kb, could co-operate to form an enhancersome for *Acan*, a functional hub that act in an additive fashion with the removal of one not impacting the others like the 5 α-goblin enhancers (Hay et al., 2016). Proximal enhancers are more active in gene regulation compared to distal regions during development and lineage determinations, *Acan* contains multiple enhancers which is not surprising as it is similar to other genes (Huang et al., 2016). What is unusual is the identification of these four new enhancers which could cause

aggrecan to be primarily expressed in chondrocytes, thus enforcing the importance ofACAN to the chondrocytes and the lethality of its loss. Their roles overlap suggesting redundancy but the -65kb and -87kb are interesting, as this stage specificity highlights questions of how important ACAN is to the hypertrophic chondrocyte? And secondly is the regulation of ACAN expression in adult chondrocytes different to that of development? Are there enhancers within the *Acan* TAD that are polycombed and repressed in development, but switched on at later stages?

To address this one of our collaborators at the Musculoskeletal biology group at the University of Newcastle, Professor David young and his student Kat Cheung have some exciting data. They are identifying human enhancers' key to the chondrocyte. In one of their studies they are assessing histone marks in human MSC and then following differentiation of the cells into chondrocytes and reassessed the histones and investigated the transcripts by RNA microarray. With their permission I was able to assess their data in regards to the *Acan* locus in terms of the chondrocyte. This data fills the gap that is missing from the ENCODE analysis. The level of *Acan* transcription by RNA array is almost absent in MSCs and but increases after 14 days differentiating into chondrocytes. The results show there is a dramatic change in the enhancer activity, there are 8 regions that are tagged with the histone marks for enhancers, 3 poised enhancers, 4 promoters and 2 strong promoters with only one region repressed in the MSC stage. In the chondrocytes there are 15 enhancers, 13 strong enhancers, 14 poised regions, 7 strong promoters and 3 promoters in the same region we examined for enhancers using ENCODE. Corresponding this to our enhancers the -87kb and +55kb is not active in MSCs, but is tagged as an enhancer in the chondrocytes, the -10kb and +26kb are enhancers that become strong, and the -65kb remains an enhancer in both stages. The -35kb was a promoter in the MSCs, but a strong promoter in chondrocytes. The most interesting result is for the +60kb that the data suggest is not active in MSC but becomes a strong enhancer in chondrocytes. This informs us that these enhancers are playing a role in the chondrocyte and they may have lineage specific roles.

Enhancers can have effects up to 1MB away such as shh enhancer that is responsible for limb, digit development (Amano et al., 2009, Sagai et al., 2005). In this study there was no scope for long range identification of enhancers that may interact with *Acan* which is an area that can be explored. We also neglected to associate these enhancers with functions of *Acan*, rather we associated them with the expression pattern similar to *Acan* and that they are in the TAD of the gene, this is emphasised by the +55kb enhancer that could be regulating *HAPLN3*.

Two interesting prospects can be put forward. *Acan* expression and turnover is critical for the normal function of joints, therefore the precise activation of *Acan* expression at decisive time points during chondrocyte maturation is important. The Examination of these enhancers under normal physiology may present minor changes but observing their roles by inactivation or in stressed situation may allow us to understand if they play a pivotal role in how skeletal and joint diseases arise. The second possibility that we put forward is that *Acan* relies on multiple enhancers to dictate the robustness of transcription. It cannot be ruled out that in the whole organism there are interactions in 3D between the regulatory elements with each other or even with the promoter to stabilise the transcriptional machinery to bring about robust expression of *Acan* (Levine et al., 2014, Vernimmen et al., 2007). This may be because *Acan* does not contain a classical TATA box or core promoter and studies have shown its reliance on SP1 and AP1 binding sites (Watanabe et al., 1995, Valhmu, 1998). One experiment that could be interesting is to determine if these enhancers can act as promoters in the absence of the core promoter to keep the production of transcripts for *Acan* in a tissue specific manner, thus providing a backup in case of mutations or loss of function in the promoter (Kowalczyk et al., 2012).

Exploration of other histone marks can detail the activity surrounding the histones such as H3K122ac and H3K64ac have been recently been identified to mark promoters and with the absence of H3K27ac marking a class of functional enhancers (Pradeepa et al., 2016). Utilising these data sets the visualisation of the *Acan* can be done and a regulatory story can be composed. For this the NCBI GEO databased contains data sets

from various experiments from across the world that can be visualised on the UCSC genome browser, each of the data sets can be annotated and layered onto the accession of interest, for this end the mm9 accession was generally used as it contained the histone modifications in the limb. The main drawback from this is that most of the data sets are created from cancer, haematopoietic or embryonic stem cells. In the Pradeepa *et al* study mentioned above they used mouse embryonic stem cells and leukaemia cell line (K562) to performed ChIP-seq and native ChIP using antibodies directed at H3K27ac, H3K4me3, H3K64ac, H3K122ac, H3K27me3 and H3K4me1 to screen for promoters and enhancers. To this end we took the data of the histone peaks for H3K64ac, H3K27ac and H3K122ac were placed into the mm9 to see if we were missing other enhancers (Figure 86). The results showed that although three regions within the *Acan* TAD has histone peaks for H3K64ac and H3K122ac, one of which is in exon, one is the -65kb and the other is the -100kb that was not analysed, suggesting they these enhancers may not fall into the active functional enhancer identified in the study (Pradeepa et al., 2016).

Likewise, ATAC-seq data in haematopoietic cells can give us an idea of the topological and 3C organisation of the chromatin around the ACAN gene using data from the ENCODE matrix from Ross Hardison lab. In multipotent, oligopotent, immortalised and lineage restricted haematopoietic cells we find that out of the 6 enhancers found in this study and the known enhancer, only the +26kb remains as open chromatin. Whereas in these cells the other intronic, -10kb,-35kb, -65kb,and -87kb remain inaccessible. This may indicate how the lineage specificity is derived in the chondrocytes and kidney/lungs for the +55kb, by changing the organisation of the chromatin. It is interesting to note that there are two other regions in these cells that are open chromatin, one is the predicted -155kb the other lies between the -35kb and -60kb that was not flagged by our ENCODE analysis. I believe that other studies can be conducted to establish changes and if these change with disease.

Not all enhancers are functionally important and a growing mass of data has generated that contradict and support the idea of enhancers having a dynamic function and are

responsive to the environment they are subjected too. One key pointis that almost all genes have more than one regulatory element and that no single *cis* or *trans* element is important without the others. A global view of haematopoietic lineage development in foetal and adult haematopoietic stem cells progenitors there are important enhancers defining each stage and a loss of any one of these enhancers delays the process. When certain enhancers are up-regulated during different differentiation stages of the cells they drive expression of the genes associated with that time point (Huang et al., 2016). The same is seen in self-renewing macrophages and embryonic stem cells but the commitment to macrophage lineages renewal is controlled by a different set of enhancers to that in the stem cells, suggesting that although certain cells in the human body share similar pathways, the control of each lineage is protected from the cross-talk of other nearby cells, this by the use of different regulatory processes found in each cell but dictated by the availability of the TFs (Soucie et al., 2016). In the well-established α-globin set of enhancers, five enhancers were identified as super-enhancers, suggesting a higher order of importance compared to other enhancers that were functionally assessed. Evaluation of each one of enhancers individually show that not one enhancer is more important, rather each enhancer acts independently and has additive function but not the synergistic effects conferred by the term super-enhancer (Hay et al., 2016). This study not only discourages the use of the term of super-enhancers, but also highlights the requirement of functionally assessment an enhancer's role. Therefore, one key aspect missing from my study is the assessment of functionality of the enhancers, do they co-operate, act independently, respond to developmental cues and are they essential to the chondrocyte?

One other factor that may contribute to an enhancer activity is if it loops the chromatin and allows the binding of activators and pioneer factors or allows interaction with the promoter directly (Chopra et al., 2012, Eun et al., 2013, Goloborodko et al., 2016b, Kagey et al., 2010, Kim et al., 2015, Krivega and Dean, 2012, Larson et al., 2012). With recent studies in super enhancers they have focused on Med1 as a marker for SEs, Med1 is a subunit of the mediator complex that interacts with the promoter and acts as an

interface for enhancers and promoters and its interactions often confer high level gene transcription. ChIP-seq data from these studies can be annotated for *Acan* to determine if any of the enhancers interact directly or via intermediates with the mediator complex (Chen and Roeder, 2011, Voss and Hager, 2014, Plaschka et al., 2015). Data from the first group to coin the name super enhancer show that there are in fact two regions that have high affinity for Med1, one is located in the 11th intron not predicted in our studies as an enhancer and may not be chondrocyte specific (Whyte et al., 2013). The other is the -65kb enhancer. This suggests that the *Acan* promoter is interacted upon at two points through enhancers and that the -65kb holds greater importance in the transcriptional regulation of the gene compared to the other enhancers (Figure 90)(Whyte et al., 2013). To determine how these interact, new Capture-C and DNA/RNA FISH would allow visualisation in fixed cells (Brackley et al., 2016, Brown et al., 2008, Brown et al., 2006). Other more advanced techniques such as modified Cas9's from different species of bacteria such as *Neisseria meningitides* and *Streptococcus thermophiles* for genome editing, can be tagged by different coloured proteins to label regions of the genome to allow multi-colour labelling of DNA and can be visualised in real time and is only limited by the power and resolution of the microscope (Ma et al., 2015a, Ma et al., 2015b, Zhang et al., 2015).

In terms of assessing these enhancers, knock out using CRISPR/cas9 or even CRISPR/cas9 panels that allow systematic identification of functional noncoding regulatory elements by targeting TFs in cells will allow an appreciation of the importance of each enhancer (Korkmaz et al., 2016). Additionally modified Cas9 and smaller cas' can allow multiple activation and deactivation of the enhancer regions without effecting the strand length or deleting and adding regions to the genome, which are previously uncontrolled variable in transgenic experiments, this is because they recognise different DNA seed sequences and do not cross-talk with each other, therefore they can be used on the same sample in one experiment (Gao et al., 2016, Konermann et al., 2015, Ran et al., 2013, Esvelt et al., 2013, Sander and Joung, 2014, Yang et al., 2013, Yang et al., 2016, Yin et al., 2014). The effects of enhancers on other regions of the genome can also be

determined if coupled with techniques such as ATAC-seq which allows the accessibility of the region to be assessed to determine if the knock-out upregulates other proximal and distal enhancers.

To understand the functional aspect of the enhancers, epigenetics and experiments with OA models such as the DMM and loading model should be carried out to determine if enhancers play a role in disease progression (Poulet, 2016, Poulet et al., 2011). Methylation is an epigenetic modification on the histones as a result of environmental stress, which can results in activation or deactivation of the region by changes the binding of *trans* factors (den Hollander et al., 2015, Reynard, 2016, Reynard et al., 2014). Differential methylation is seen in OA patients and differences are seen in hip and knee OA, the arcOGEN GWAS and other methylation studies have found susceptibility loci for OA in *GNL3, ASTN2* and *CHST11* (arcOGEN, 2012, Nakajima et al., 2010, Kerkhof et al., 2010, Gee et al., 2014). With help from Dr Louise Reynard in the osteoarthritis genetics group at the University of Newcastle, we were able to analyse the methylation data for knee and hip OA to determine if the *Acan* enhancers are differentially methylated. The analysed can data found within the appendix was reproduced with permission. Data was not available for the -87kb enhancer due to no probe coverage and the only region to be methylated is the -65kb which differentially methylated when compared healthy hip to OA hip in patients (Rushton et al., 2014). The exact role of this methylation, requires validation as methylation is not the only mechanism by which enhancers can be changed or modulatd in disease. To examine the role of the enhancers in a condition like OA, we used an *in vitro* analysis, using the ATDC5 cells. I transfected the cells with luciferase reporter constructs with the enhancers upstream of a minimal promoter, we opted against using the HSP68/*LacZ* vector *in vitro* and using a Dual-light assay that detects luciferase and β-gal activity, due to the fact the HSP68, although silent in mice is not silent *in vitro*. After verifying that the enhancers were able to express *in vitro* using the T/C28-I2 and NIH3T3 cells, we transfected the ATDC5, after differentiation, with the enhancer constructs and then added hOSM or IL-1β. Here the -35kb element was poorly expressed despite its activity *in vivo*, the -65kb was also unable to express as strongly as

the other enhancers. The -87kb appears to be chondrocyte specific, as in fibroblast the activity is greatly reduced. In the ATDC5 cells there were no significant effects of hOSM or IL-1β on the expression of the +26kb and -10kb. The -35kb increased in expression with IL-1β treatment, but this was not significant. The effects on -65kb by IL-1β hints that the -65kb is a responsive element and may be critical to increasing anabolic activity, the -87kb on the other hand responds to hOSM that down regulates its expression. OSM and IL-1β signal through different pathways, IL-1β works by signalling through the TRAF-6/TAK-1/MKK-4/JNK-2 to promote aggrecanolysis by increasing the expression of ADAMTS-5, MMP1 and MMP13 in OA as well as reducing ACAN expression and downregulating *Acan* mRNA (Ismail et al., 2015a, Ismail et al., 2015b, Radons et al., 2006). OSM signals through a JNK-2 independent pathway via the JAK/STAT and down regulates aggrecanase activity, OSM can also act synergistically with IL-1β (Rogerson et al., 2010, Barksby et al., 2006). The -65kb may be the developmentally important enhancer and as it is methylated in hip OA, it may point to a switch in OA cartilage to attempt to synthesise more matrix but due to methylation this is hindered. The -87kb appears to be important for the chondrocytes expression of Acan and the down regulation by OSM and Il-1β may be indication at its role in turnover or a lack of turnover. This may suggests that there are defined roles for these enhancers in adult tissue. This may be the turnover of ACAN but this may be limited as the half-life of ACAN in adult cartilage is around 3.4 years (Maroudas et al., 1998). Therefore, an area for further work would be to define the role of these enhancers in adult cartilage. These enhancers may play a role in disease progress such as OA, although the data is very preliminary and it is not unusual for SNPs in noncoding elements predisposing a person to a disease Enhancers may be the regions of DNA that are blocked or up-regulated in diseases such as OA, that adds to the severity of the disease or prevents a response for the cell to rebuild its ECM. One enhancer may not be able to change a disease state, it may require other enhancers or other factors, such as targeting not just the *Acan* enhancer but the enhancers in collagen 2 and down-regulating the enhancers in enzymes such as MMP13. .

With the available data from ChIP-seq, ATAC-seq, ChIA-PET and combined with our analysis using EMSA and transgenic mice, these together build a picture of the developmental regulation of *Acan*, although a view of what may happen in the adult remains to be examined. From this a picture can be painted of the chromatin landscape of the *Acan* TAD, the 3D structure of this "hub" can be determined with new technology and analytic tools being developed such as the PETModule and the 4D genome (Zhao et al., 2016, Teng et al., 2015). From the 4D genome approach looking at Hi-C of the IMR-90 human lung cells there is no interactions between the +26kb enhancer with any of the enhancers of Acan, there is however interactions with other upstream regions of *Acan* and within the *HAPLN3* gene. The promoter (taken as chr15:89,346,048-89,347,590) interacts with 34 regions, but of the enhancers the -10kb, -35kb and -87kb, the -35kb interacts with the -87kb and -65kb and the -87kb only interacts with the promoter and -35kb (Figure 91). This data may be biased to the lungs and in chondrocytes this may be completely different, but it gives us invaluable insight and the interactions with -65kb and -35kb may strengthen the idea of the -35kb playing the role of a promoter.

5.9. Using Acan Cre as a tool for further research

ACAN is express more steadily in adult cartilage than Col2a1 and is found both in the growth plate, articular cartilage and IVD but *Col2a1* gradually decreases with age (Chambers et al., 2002, Neidlinger-Wilke et al., 2006, Chen et al., 2007). Therefore, there are advantage to using an *Acan* cre over a *Col2a1* and allows control other tissue of activation and temporal control over the Cre's activity (Ovchinnikov et al., 2000, Nakamura et al., 2006). There is an established *Acan* Cre that was a knock in into the 3'UTR of *Acan* which allows analysis of floxed genes *in vivo* during development and adulthood (Henry et al., 2009). However, there are some draw backs as chondrocytes below the tide mark and in the deeper layers of the meniscus were not activated. Our group has also created a *Acan*CreER^{T2} using the known A1 enhancer sequence in four

time tandem repeat, cloned upstream of the *Col2a1* proximal promoter that was different from the knock-in cre as it the lacked tendon expression, therefore, we now have multiple transgenic mice models that can help us understand cartilage disease by using them in over-expressing or knock down of a gene specifically in the chondrocyte (Cascio et al., 2014).

The drawback of other cre models are the non-chondrogenic expression in development which warrants a specific *Acan* cre. However, the reason to add the -35kb enhancer upstream of the human *Acan* minimal promote to the arsenal of *Acan* cres was the strength of the enhancer and and its activity in both development and adulthood.

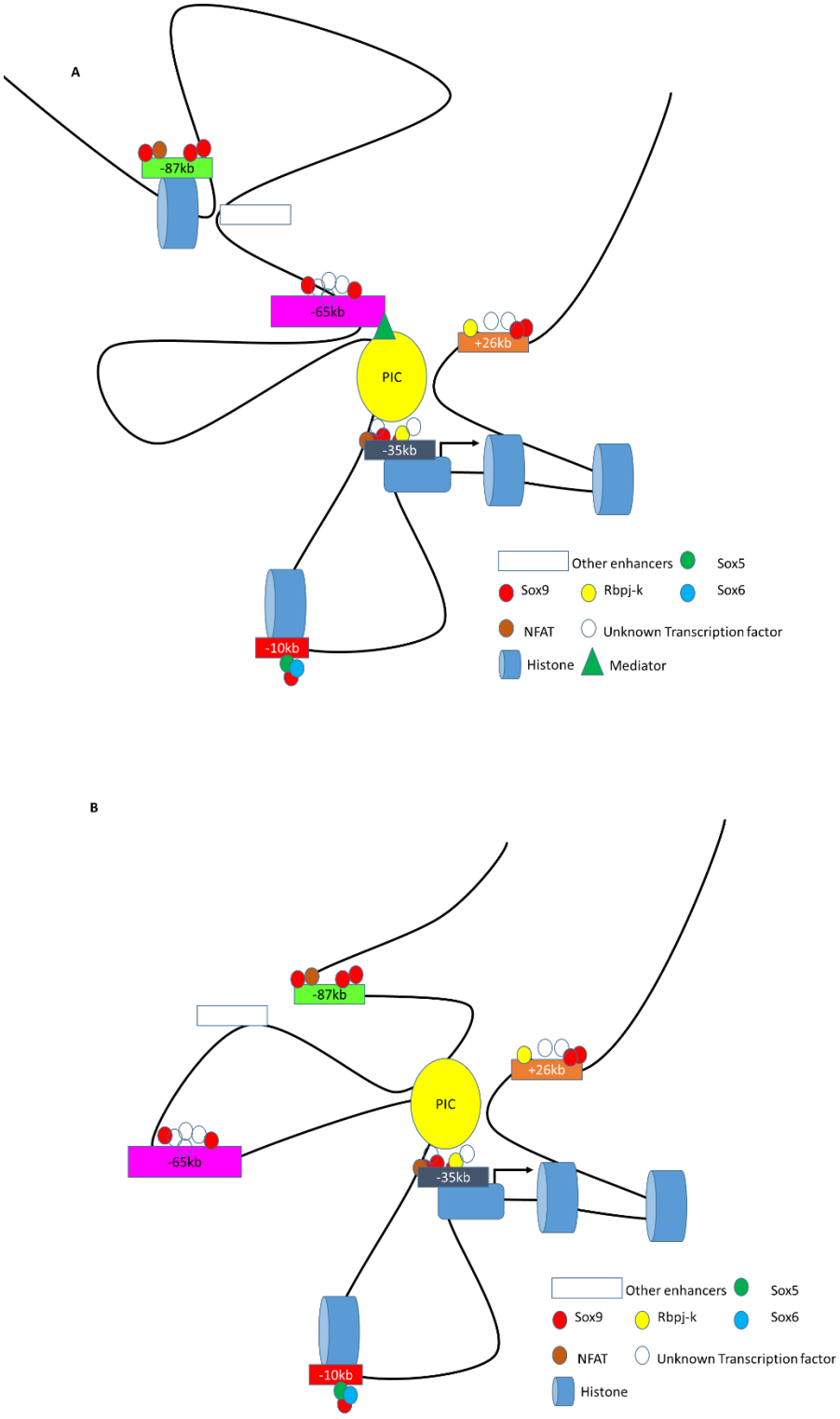
Additionally, when compared to the known A1 enhancer and the knock in cre, the -35kb driving the human promoter may be of a different lineage, or subset of chondrocytes, so the use of a different *Acan* cre can determine cell fate by mapping the expression after pulse chasing (Carlone, 2016, Wuidart et al., 2016, Vorhagen et al., 2015). This will shine a light on the identity of the chondrocytes and the different fate in which they may take and allow us to differentiate chondrocytes from for example the superficial articular chondrocyte from the rib chondrocytes.

5.10. Conclusions

Multiple enhancers were identified in the *Acan* locus that expressin chondrocytes as well as other intronic *cis*-acting elements that express elsewhere. The role of each enhancer remains to be assessed and their gene targets will need to be verified. Sox9 regulates the -35kb enhancer *in vivo* and *in vitro* experiments show that the TF binds to all the chondrocyte expressed enhancers. The role of Sox9 may be structural and an activator in development, but its role outside of development in regards to these enhancers remains to be explored as well as the 3D chromatin landscape. These enhancers may act collectively or as an additive manner. Developmental interactions may be only required to allow robust expression of ACAN, but in adult tissues the enhancers act in a different capacity or to react to stress this is depicted in Figure 83. The developmental and adult landscape are different and it may be that the -87kb is the

most important in preserving *Acan* transcription in adulthood but the -65kb is essential for gene activation in development, images depicting the working models for the enhancers are outlined in Figure 83. The role of these enhancer in disease is unclear and their response may play critical roles in disease progression, although our preliminary data suggest they may be involved in the response cytokines that have a role in OA disease progression, their actual function in disease state may be minimal but requires further investigation. The pChu94 mouse is an inducible chondrocyte specific model that can be used to interrogate the loss of genes in chondrocytes or to assess the role of certain pathways in both development and adult cartilage.

Figure 83: Schematic of possible enhancer interactions based on known data



There are many possibilities for how these enhancers interact. They may not help each other but rather they may act as independent entities. A collation of the data we know it seems more plausible the -65kb is the key enhancer in keeping Acan transcription along with a region up stream that has not been identified in this study. The -65kb interacts with the mediator to drive the expression of Acan in a timely manner during development (A). The -10kb may loop the DNA to stabilise the transcriptional loops at the proximal promoter. The -35kb may acts as a promoter for the chondrocyte with the +26kb bringing downstream elements closer to the hub, again in development the -87kb is looped out. These 5 enhancers are key to the lineage determination of the chondrocyte, as they are the only regions to interact with Sox9 at E12, they act in an additive manner to stabilise the entire complex with fail safes to ensure the chondrocyte expresses Acan to keep the integrity of the cartilage. As the MSC commit to the chondrocyte each of the enhancer become active and the -87kb is switched on or even looped in to interact with the promoter, as the chondrocyte undergoes it lineage towards the articular cartilage the -87kb plays a more important role. These roles are yet to be defined.

5.11. Future direction

Functionally assessing enhancers and determining their functional relevance is pinnacle as it may reveal the role each enhancer works (Kvon, 2015, Hay et al., 2016). Determining the role of enhancers in adulthood and maintenance of the genomic landscape, appears to be important and differences in development and adult transcriptional regulation can be examined with the advancing field of molecular biology and the embracing genome editing tools. Establishing the role enhancers' play on a global genomic scale can establish the role regulatory elements found in cartilage give rise to the entire join and the tissues that are found there, there may be regulatory elements that we are missing altogether and need to be considered as dynamic elements are forever changing. Targeting the -65kb and-87kb may be a possible area for OA treatment but requires further research into their function in disease and how

impactful they are if targeted. A major aim of future research is to understand the role of the enhancers and how they communicate to bring about co-ordinated development and maintaince of a whole organism, if this is dysregulated in diseases or can be targeted.

6. REFERENCES

Abbasi, A. A., Paparidis, Z., Malik, S., Bangs, F., Schmidt, A., Koch, S., Lopez-Rios, J. & Grzeschik, K. H. 2010. Human intronic enhancers control distinct sub-domains of Gli3 expression during mouse CNS and limb development. *BMC Dev Biol,* 10**,** 44.

Abrahamson, D. R. 2009. Development of kidney glomerular endothelial cells and their role in basement membrane assembly. *Organogenesis,* 5**,** 275-287.

Ahmad, K. & Henikoff, S. 2002. The Histone Variant H3.3 Marks Active Chromatin by Replication-Independent Nucleosome Assembly. *Molecular Cell,* 9**,** 1191-1200.

Akiyama, H. 2008. Control of chondrogenesis by the transcription factor Sox9. *Mod Rheumatol,* 18**,** 213-9.

Akiyama, H., Chaboissier, M. C., Martin, J. F., Schedl, A. & De Crombrugghe, B. 2002. The transcription factor Sox9 has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of Sox5 and Sox6. *Genes Dev,* 16**,** 2813-28.

Akiyama, H., Kim, J. E., Nakashima, K., Balmes, G., Iwai, N., Deng, J. M., Zhang, Z., Martin, J. F., Behringer, R. R., Nakamura, T. & De Crombrugghe, B. 2005. Osteo-chondroprogenitor cells are derived from Sox9 expressing precursors. *Proc Natl Acad Sci U S A,* 102**,** 14665-70.

Akiyama, H., Lyons, J. P., Mori-Akiyama, Y., Yang, X., Zhang, R., Zhang, Z., Deng, J. M., Taketo, M. M., Nakamura, T., Behringer, R. R., Mccrea, P. D. & De Crombrugghe, B. 2004. Interactions between Sox9 and beta-catenin control chondrocyte differentiation. *Genes Dev,* 18**,** 1072-87.

Akkiraju, H. & Nohe, A. 2015. Role of Chondrocytes in Cartilage Formation, Progression of Osteoarthritis and Cartilage Regeneration. *Journal of developmental biology,* 3**,** 177-192.

Albro, M. B., Banerjee, R. E., Li, R., Oungoulian, S. R., Chen, B., Del Palomar, A. P., Hung, C. T. & Ateshian, G. A. 2011. Dynamic loading of immature epiphyseal cartilage pumps nutrients out of vascular canals. *J Biomech,* 44**,** 1654-9.

Ali, T., Renkawitz, R. & Bartkuhn, M. 2016. Insulators and domains of gene expression. *Current Opinion in Genetics & Development,* 37**,** 17-26.

Alvarez, J., Costales, L., Serra, R., Balbin, M. & Lopez, J. M. 2005. Expression patterns of matrix metalloproteinases and vascular endothelial growth factor during epiphyseal ossification. *J Bone Miner Res,* 20**,** 1011-21.

Amano, T., Sagai, T., Tanabe, H., Mizushina, Y., Nakazawa, H. & Shiroishi, T. 2009. Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. *Dev Cell,* 16**,** 47-57.

Anson, D. S. & Limberis, M. 2004. An improved β-galactosidase reporter gene. *Journal of Biotechnology,* 108**,** 17-30.

Antoniv, T. T., De Val, S., Wells, D., Denton, C. P., Rabe, C., De Crombrugghe, B., Ramirez, F. & Bou-Gharios, G. 2001. Characterization of an evolutionarily conserved far-upstream enhancer in the human alpha 2(I) collagen (COL1A2) gene. *J Biol Chem,* 276**,** 21754-64.

Anwar, S., Minhas, R., Ali, S., Lambert, N., Kawakami, Y., Elgar, G., Azam, S. S. & Abbasi, A. A. 2015. Identification and functional characterization of novel transcriptional enhancers involved in regulating human GLI3 expression during early development. *Dev Growth Differ,* 57**,** 570-80.

Arciniegas, E., Neves, C. Y., Candelle, D. & Parada, D. 2004. Differential versican isoforms and aggrecan expression in the chicken embryo aorta. *Anat Rec A Discov Mol Cell Evol Biol,* 279**,** 592-600.

Arcogen, U. 2012. Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. *Lancet,* 380.

Arnold, M. A., Kim, Y., Czubryt, M. P., Phan, D., Mcanally, J., Qi, X., Shelton, J. M., Richardson, J. A., Bassel-Duby, R. & Olson, E. N. 2007. MEF2C transcription factor controls chondrocyte hypertrophy and bone development. *Dev Cell,* 12**,** 377-89.

Aspberg, A. 2012. The different roles of aggrecan interaction domains. *J Histochem Cytochem,* 60**,** 987-96.

Atsumi, T., Ikawa, Y., Miwa, Y. & Kimata, K. 1990. A chondrogenic cell line derived from a differentiating culture of AT805 teratocarcinoma cells. *Cell Differentiation and Development,* 30**,** 109-116.

Aza-Carmona, M., Barca-Tierno, V., Hisado-Oliva, A., Belinchon, A., Gorbenko-Del Blanco, D., Rodriguez, J. I., Benito-Sanz, S., Campos-Barros, A. & Heath, K. E. 2014. NPPB and ACAN, two novel SHOX2 transcription targets implicated in skeletal development. *PLoS One,* 9**,** e83104.

Aza-Carmona, M., Shears, D. J., Yuste-Checa, P., Barca-Tierno, V., Hisado-Oliva, A., Belinchon, A., Benito-Sanz, S., Rodriguez, J. I., Argente, J., Campos-Barros, A., Scambler, P. J. & Heath, K. E. 2011. SHOX interacts with the chondrogenic transcription factors SOX5 and SOX6 to activate the aggrecan enhancer. *Hum Mol Genet,* 20**,** 1547-59.

Bachmair, A., Finley, D. & Varshavsky, A. 1986. In vivo half-life of a protein is a function of its amino-terminal residue. *Science,* 234**,** 179-186.

Bahney, C. S., Hu, D. P., Taylor, A. J., Ferro, F., Britz, H. M., Hallgrimsson, B., Johnstone, B., Miclau, T. & Marcucio, R. S. 2014. Stem cell-derived endochondral cartilage stimulates bone healing by tissue transformation. *J Bone Miner Res,* 29**,** 1269-82.

Banerji, J., Rusconi, S. & Schaffner, W. 1981. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell,* 27**,** 299-308.

Barksby, H. E., Hui, W., Wappler, I., Peters, H. H., Milner, J. M., Richards, C. D., Cawston, T. E. & Rowan, A. D. 2006. Interleukin-1 in combination with oncostatin M up-regulates multiple genes in chondrocytes: implications for cartilage destruction and repair. *Arthritis Rheum,* 54**,** 540-50.

Barna, M. & Niswander, L. 2007. Visualization of cartilage formation: insight into cellular properties of skeletal progenitors and chondrodysplasia syndromes. *Dev Cell,* 12**,** 931-41.

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I. & Zhao, K. 2007. High-resolution profiling of histone methylations in the human genome. *Cell,* 129**,** 823-37.

Becker, P. W., Sacilotto, N., Nornes, S., Neal, A., Thomas, M. O., Liu, K., Preece, C., Ratnayaka, I., Davies, B., Bou-Gharios, G. & De Val, S. 2016. An Intronic Flk1 Enhancer Directs Arterial-Specific Expression via RBPJ-Mediated Venous Repression. *Arterioscler Thromb Vasc Biol,* 36**,** 1209-19.

Beier, F., Vornehm, S., Poschl, E., Von Der Mark, K. & Lammi, M. J. 1997. Localization of silencer and enhancer elements in the human type X collagen gene. *J Cell Biochem,* 66**,** 210-8.

Beiser, K. U., Glaser, A., Kleinschmidt, K., Scholl, I., Roth, R., Li, L., Gretz, N., Mechtersheimer, G., Karperien, M., Marchini, A., Richter, W. & Rappold, G. A. 2014. Identification of Novel SHOX Target Genes in the Developing Limb Using a Transgenic Mouse Model. *PLoS One,* 9**,** e98543.

Belluoccio, D., Wilson, R., Thornton, D. J., Wallis, T. P., Gorman, J. J. & Bateman, J. F. 2006. Proteomic analysis of mouse growth plate cartilage. *Proteomics,* 6**,** 6549-53.

Belton, J. M., Mccord, R. P., Gibcus, J. H., Naumova, N., Zhan, Y. & Dekker, J. 2012. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods,* 58**,** 268-76.

Benazet, J. D., Pignatti, E., Nugent, A., Unal, E., Laurent, F. & Zeller, R. 2012. Smad4 is required to induce digit ray primordia and to initiate the aggregation and differentiation of chondrogenic progenitors in mouse limb buds. *Development,* 139**,** 4250-60.

Bender, M. A., Byron, R., Ragoczy, T., Telling, A., Bulger, M. & Groudine, M. 2006. Flanking HS-62.5 and 3' HS1, and regions upstream of the LCR, are not required for β-globin transcription. *Blood,* 108**,** 1395-1401.

Bender, M. A., Ragoczy, T., Lee, J., Byron, R., Telling, A., Dean, A. & Groudine, M. 2012. The hypersensitive sites of the murine beta-globin locus control region act independently to affect nuclear localization and transcriptional elongation. *Blood,* 119**,** 3820-7.

Berendsen, A. D. & Olsen, B. R. 2015. Bone development. *Bone,* 80**,** 14-18.

Beura, L. K., Hamilton, S. E., Bi, K., Schenkel, J. M., Odumade, O. A., Casey, K. A., Thompson, E. A., Fraser, K. A., Rosato, P. C., Filali-Mouhim, A., Sekaly, R. P., Jenkins, M. K., Vezys, V., Haining, W. N., Jameson, S. C. & Masopust, D. 2016. Normalizing the environment recapitulates adult human immune traits in laboratory mice. *Nature,* 532**,** 512-516.

Bi, W., Deng, J. M., Zhang, Z., Behringer, R. R. & De Crombrugghe, B. 1999. Sox9 is required for cartilage formation. *Nat Genet,* 22**,** 85-9.

Bi, W., Huang, W., Whitworth, D. J., Deng, J. M., Zhang, Z., Behringer, R. R. & De Crombrugghe, B. 2001. Haploinsufficiency of Sox9 results in defective cartilage primordia and premature skeletal mineralization. *Proc Natl Acad Sci U S A,* 98**,** 6698-703.

Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigo, R., Gingeras, T. R., Margulies, E. H., Weng, Z., Snyder, M., Dermitzakis, E. T., Thurman, R. E., Kuehn, M. S., Taylor, C. M., Neph, S., Koch, C. M., Asthana, S., Malhotra, A., Adzhubei, I., Greenbaum, J. A., Andrews, R. M., Flicek, P., Boyle, P. J., Cao, H., Carter, N. P., Clelland, G. K., Davis, S., Day, N., Dhami, P., Dillon, S. C., Dorschner, M. O., Fiegler, H., Giresi, P. G., Goldy, J., Hawrylycz, M., Haydock, A., Humbert, R., James, K. D., Johnson, B. E., Johnson, E. M., Frum, T. T., Rosenzweig, E. R., Karnani, N., Lee, K., Lefebvre, G. C., Navas, P. A., Neri, F., Parker, S. C., Sabo, P. J., Sandstrom, R., Shafer, A., Vetrie, D., Weaver, M., Wilcox, S., Yu, M., Collins, F. S., Dekker, J., Lieb, J. D., Tullius, T. D., Crawford, G. E., Sunyaev, S., Noble, W. S., Dunham, I., Denoeud, F., Reymond, A., Kapranov, P., Rozowsky, J., Zheng, D., Castelo, R., Frankish, A., Harrow, J., Ghosh, S., Sandelin, A., Hofacker, I. L., Baertsch, R., Keefe, D., Dike, S., Cheng, J., Hirsch, H. A., Sekinger, E. A., Lagarde, J., Abril, J. F., Shahab, A., Flamm, C., Fried, C., Hackermuller, J., Hertel, J., Lindemeyer, M., Missal, K., Tanzer, A., Washietl, S., Korbel, J., Emanuelsson, O., Pedersen, J. S., Holroyd, N., Taylor, R., Swarbreck, D., Matthews, N., Dickson, M. C., Thomas, D. J., Weirauch, M. T., Gilbert, J., et al. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature,* 447**,** 799-816.

Blattler, A., Yao, L., Witt, H., Guo, Y., Nicolet, C. M., Berman, B. P. & Farnham, P. J. 2014. Global loss of DNA methylation uncovers intronic enhancers in genes showing expression changes. *Genome Biol,* 15**,** 469.

Blumer, M. J. F., Longato, S., Richter, E., Pérez, M. T., Konakci, K. Z. & Fritsch, H. 2005. The role of cartilage canals in endochondral and perichondral bone formation: are there similarities between these two processes? *Journal of Anatomy,* 206**,** 359-372.

Bobick, B. E. & Cobb, J. 2012. Shox2 regulates progression through chondrogenesis in the mouse proximal limb. *J Cell Sci,* 125**,** 6071-83.

Boeger, H., Griesenbeck, J., Strattan, J. S. & Kornberg, R. D. 2003. Nucleosomes Unfold Completely at a Transcriptionally Active Promoter. *Molecular Cell,* 11**,** 1587-1598.

Bothma, Jacques p., Magliocco, J. & Levine, M. 2011. The Snail Repressor Inhibits Release, Not Elongation, of Paused Pol II in the Drosophila Embryo. *Current Biology,* 21**,** 1571-1577.

Bou-Gharios, G., Garrett, L. A., Rossert, J., Niederreither, K., Eberspaecher, H., Smith, C., Black, C. & Crombrugghe, B. 1996. A potent far-upstream enhancer in the mouse pro alpha 2(I) collagen gene regulates expression of reporter genes in transgenic mice. *J Cell Biol,* 134**,** 1333-44.

Brackley, C. A., Brown, J. M., Waithe, D., Babbs, C., Davies, J., Hughes, J. R., Buckle, V. J. & Marenduzzo, D. 2016. Predicting the three-dimensional folding of cis-regulatory regions in mammalian genomes using bioinformatic data and polymer models. *Genome Biol,* 17**,** 59.

Bridgewater, L. C., Lefebvre, V. & De Crombrugghe, B. 1998. Chondrocyte-specific enhancer elements in the Col11a2 gene resemble the Col2a1 tissue-specific enhancer. *J Biol Chem,* 273**,** 14998-5006.

Bridgewater, L. C., Walker, M. D., Miller, G. C., Ellison, T. A., Holsinger, L. D., Potter, J. L., Jackson, T. L., Chen, R. K., Winkel, V. L., Zhang, Z., Mckinney, S. & De Crombrugghe, B. 2003. Adjacent DNA sequences modulate Sox9 transcriptional

activation at paired Sox sites in three chondrocyte-specific enhancer elements. *Nucleic Acids Res,* 31**,** 1541-53.

Briot, A., Jaroszewicz, A., Warren, Carmen m., Lu, J., Touma, M., Rudat, C., Hofmann, Jennifer j., Airik, R., Weinmaster, G., Lyons, K., Wang, Y., Kispert, A., Pellegrini, M. & Iruela-Arispe, M. L. 2014. Repression of Sox9 by Jag1 Is Continuously Required to Suppress the Default Chondrogenic Fate of Vascular Smooth Muscle Cells. *Developmental Cell,* 31**,** 707-721.

Brocard, J., Warot, X., Wendling, O., Messaddeq, N., Vonesch, J. L., Chambon, P. & Metzger, D. 1997. Spatio-temporally controlled site-specific somatic mutagenesis in the mouse. *Proc Natl Acad Sci U S A,* 94**,** 14559-63.

Brown, J. M., Green, J., Das Neves, R. P., Wallace, H. A., Smith, A. J., Hughes, J., Gray, N., Taylor, S., Wood, W. G., Higgs, D. R., Iborra, F. J. & Buckle, V. J. 2008. Association between active genes occurs at nuclear speckles and is modulated by chromatin environment. *J Cell Biol,* 182**,** 1083-97.

Brown, J. M., Leach, J., Reittie, J. E., Atzberger, A., Lee-Prudhoe, J., Wood, W. G., Higgs, D. R., Iborra, F. J. & Buckle, V. J. 2006. Coregulated human globin genes are frequently in spatial proximity when active. *J Cell Biol,* 172**,** 177-87.

Bulger, M. & Groudine, M. 2010. Enhancers: the abundance and function of regulatory sequences beyond promoters. *Dev Biol,* 339**,** 250-7.

Burke, T. W. & Kadonaga, J. T. 1996. Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes & Development,* 10**,** 711-724.

Burke, T. W. & Kadonaga, J. T. 1997. The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila. *Genes & Development,* 11**,** 3020-3031.

Burr, D. B. & Gallant, M. A. 2012. Bone remodelling in osteoarthritis. *Nat Rev Rheumatol,* 8**,** 665-673.

Cannavò, E., Khoueiry, P., Garfield, David a., Geeleher, P., Zichner, T., Gustafson, E. H., Ciglar, L., Korbel, Jan o. & Furlong, Eileen e. M. 2016. Shadow Enhancers Are

Pervasive Features of Developmental Regulatory Networks. *Current Biology,* 26**,** 38-51.

Cao, L., Zhang, Y. & Yang, B. B. 1998. Expression of the G1 domain of aggrecan interferes with chondrocyte attachment and adhesion. *Matrix Biol,* 17**,** 379-92.

Caporali, E. H., Kuykendall, T. & Stewart, M. C. 2015. Complete sequencing and characterization of equine aggrecan. *Vet Comp Orthop Traumatol,* 28**,** 79-87.

Carey, M., Lin, Y. S., Green, M. R. & Ptashne, M. 1990. A mechanism for synergistic activation of a mammalian gene by GAL4 derivatives. *Nature,* 345**,** 361-4.

Carlone, D. L. 2016. Identifying Adult Stem Cells Using Cre-Mediated Lineage Tracing. *Curr Protoc Stem Cell Biol,* 36**,** 5a.2.1-18.

Carroll, K. J., Makarewich, C. A., Mcanally, J., Anderson, D. M., Zentilin, L., Liu, N., Giacca, M., Bassel-Duby, R. & Olson, E. N. 2016. A mouse model for adult cardiac-specific gene deletion with CRISPR/Cas9. *Proceedings of the National Academy of Sciences,* 113**,** 338-343.

Cascio, L. L., Liu, K., Nakamura, H., Chu, G., Lim, N. H., Chanalaris, A., Saklatvala, J., Nagase, H. & Bou-Gharios, G. 2014. Generation of a mouse line harboring a bi-transgene expressing luciferase and tamoxifen-activatable creERT2 recombinase in cartilage. *genesis,* 52**,** 110-119.

Catalina Cabrera, L., Mcnabb, B. R., Woods, S. E., Cartoceti, A. N. & Busch, R. C. 2016. Hydrops associated with chondrodysplasia of the fetus in a miniature Scottish Highland cow. *J Am Vet Med Assoc,* 248**,** 552-6.

Caterina, J. J., Ciavatta, D. J., Donze, D., Behringer, R. R. & Townes, T. M. 1994. Multiple elements in human beta-globin locus control region 5' HS 2 are involved in enhancer activity and position-independent, transgene expression. *Nucleic Acids Research,* 22**,** 1006-1011.

Cavanagh, J. A., Tammen, I., Windsor, P. A., Bateman, J. F., Savarirayan, R., Nicholas, F. W. & Raadsma, H. W. 2007. Bulldog dwarfism in Dexter cattle is caused by mutations in ACAN. *Mamm Genome,* 18**,** 808-14.

Chamberland, A., Wang, E., Jones, A. R., Collins-Racie, L. A., Lavallie, E. R., Huang, Y., Liu, L., Morris, E. A., Flannery, C. R. & Yang, Z. 2009. Identification of a novel HtrA1-susceptible cleavage site in human aggrecan: evidence for the involvement of HtrA1 in aggrecan proteolysis in vivo. *J Biol Chem,* 284**,** 27352-9.

Chambers, M. G., Kuffner, T., Cowan, S. K., Cheah, K. S. & Mason, R. M. 2002. Expression of collagen and aggrecan genes in normal and osteoarthritic murine knee joints. *Osteoarthritis Cartilage,* 10**,** 51-61.

Chen, H., Lowrey, C. H. & Stamatoyannopoulos, G. 1997. Analysis of enhancer function of the HS-40 core sequence of the human alpha-globin cluster. *Nucleic Acids Research,* 25**,** 2917-2922.

Chen, K., Chen, Z., Wu, D., Zhang, L., Lin, X., Su, J., Rodriguez, B., Xi, Y., Xia, Z., Chen, X., Shi, X., Wang, Q. & Li, W. 2015. Broad H3K4me3 is associated with increased transcription elongation and enhancer activity at tumor-suppressor genes. *Nat Genet,* 47**,** 1149-1157.

Chen, L., Fink, T., Ebbesen, P. & Zachar, V. 2006. Temporal transcriptome of mouse ATDC5 chondroprogenitors differentiating under hypoxic conditions. *Experimental Cell Research,* 312**,** 1727-1744.

Chen, L., Fink, T., Zhang, X.-Y., Ebbesen, P. & Zachar, V. 2005. Quantitative transcriptional profiling of ATDC5 mouse progenitor cells during chondrogenesis. *Differentiation,* 73**,** 350-363.

Chen, L. & Widom, J. 2005. Mechanism of Transcriptional Silencing in Yeast. *Cell,* 120**,** 37-48.

Chen, M., Lichtler, A. C., Sheu, T. J., Xie, C., Zhang, X., O'keefe, R. J. & Chen, D. 2007. Generation of a transgenic mouse model with chondrocyte-specific and tamoxifen-inducible expression of Cre recombinase. *Genesis,* 45**,** 44-50.

Chen, S., Tao, J., Bae, Y., Jiang, M. M., Bertin, T., Chen, Y., Yang, T. & Lee, B. 2013. Notch gain of function inhibits chondrocyte differentiation via Rbpj-dependent suppression of Sox9. *J Bone Miner Res,* 28**,** 649-59.

Chen, W. & Roeder, R. G. 2011. Mediator-dependent nuclear receptor function. *Semin Cell Dev Biol,* 22**,** 749-58.

Chopra, V. S., Kong, N. & Levine, M. 2012. Transcriptional repression via antilooping in the Drosophila embryo. *Proceedings of the National Academy of Sciences,* 109**,** 9460-9464.

Chu, C., Zhang, Qiangfeng c., Da rocha, Simão t., Flynn, Ryan a., Bharadwaj, M., Calabrese, J. M., Magnuson, T., Heard, E. & Chang, Howard y. 2015a. Systematic Discovery of Xist RNA Binding Proteins. *Cell,* 161**,** 404-416.

Chu, V. T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K. & Kuhn, R. 2015b. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat Biotechnol*.

Clevers, H. 2006. Wnt/beta-catenin signaling in development and disease. *Cell,* 127**,** 469-80.

Colombier, P., Clouet, J., Hamel, O., Lescaudron, L. & Guicheux, J. 2014. The lumbar intervertebral disc: from embryonic development to degeneration. *Joint Bone Spine,* 81**,** 125-9.

Conaway, R. C. & Conaway, J. W. 1993. General Initiation Factors for RNA Polymerase II. *Annual Review of Biochemistry,* 62**,** 161-190.

Cooper, K. L., Oh, S., Sung, Y., Dasari, R. R., Kirschner, M. W. & Tabin, C. J. 2013. Multiple phases of chondrocyte enlargement underlie differences in skeletal proportions. *Nature,* 495**,** 375-8.

Cosgrove, M. S., Boeke, J. D. & Wolberger, C. 2004. Regulated nucleosome mobility and the histone code. *Nat Struct Mol Biol,* 11**,** 1037-43.

Coustry, F., Oh, C.-D., Hattori, T., Maity, S. N., De Crombrugghe, B. & Yasuda, H. 2010. The dimerization domain of SOX9 is required for transcription activation of a chondrocyte-specific chromatin DNA template. *Nucleic Acids Research,* 38**,** 6018-6028.

Cromar, G., Wong, K.-C., Loughran, N., On, T., Song, H., Xiong, X., Zhang, Z. & Parkinson, J. 2014. New Tricks for “Old” Domains: How Novel Architectures and

Promiscuous Hubs Contributed to the Organization and Evolution of the ECM. *Genome Biology and Evolution,* 6**,** 2897-2917.

Cserjesi, P., Brown, D., Ligon, K. L., Lyons, G. E., Copeland, N. G., Gilbert, D. J., Jenkins, N. A. & Olson, E. N. 1995. Scleraxis: a basic helix-loop-helix protein that prefigures skeletal formation during mouse embryogenesis. *Development,* 121**,** 1099-1110.

Daoud, G., Kempf, H., Kumar, D., Kozhemyakina, E., Holowacz, T., Kim, D. W., Ionescu, A. & Lassar, A. B. 2014. BMP-mediated induction of GATA4/5/6 blocks somitic responsiveness to SHH. *Development,* 141**,** 3978-87.

David, G., Lories, V., Heremans, A., Van Der Schueren, B., Cassiman, J. J. & Van Den Berghe, H. 1989. Membrane-associated chondroitin sulfate proteoglycans of human lung fibroblasts. *J Cell Biol,* 108**,** 1165-73.

De Val, S., Ponticos, M., Antoniv, T. T., Wells, D. J., Abraham, D., Partridge, T. & Bou-Gharios, G. 2002. Identification of the key regions within the mouse pro-alpha 2(I) collagen gene far-upstream enhancer. *J Biol Chem,* 277**,** 9286-92.

Deaton, A. M. & Bird, A. 2011. CpG islands and the regulation of transcription. *Genes Dev,* 25**,** 1010-22.

Decker, R. S., Koyama, E., Enomoto-Iwamoto, M., Maye, P., Rowe, D., Zhu, S., Schultz, P. G. & Pacifici, M. 2014a. Mouse limb skeletal growth and synovial joint development are coordinately enhanced by Kartogenin. *Dev Biol,* 395**,** 255-67.

Decker, R. S., Koyama, E. & Pacifici, M. 2014b. Genesis and morphogenesis of limb synovial joints and articular cartilage. *Matrix Biol*.

Decker, R. S., Koyama, E. & Pacifici, M. 2015. Articular Cartilage: Structural and Developmental Intricacies and Questions. *Curr Osteoporos Rep*.

Dejardin, J. 2015. Switching between Epigenetic States at Pericentromeric Heterochromatin. *Trends Genet,* 31**,** 661-72.

Dekker, J. & Mirny, L. 2016. The 3D Genome as Moderator of Chromosomal Communication. *Cell,* 164**,** 1110-1121.

Den Hollander, W., Ramos, Y. F., Bomer, N., Elzinga, S., Van Der Breggen, R., Lakenberg, N., De Dijcker, W. J., Suchiman, H. E., Duijnisveld, B. J., Houwing-Duistermaat, J.

J., Slagboom, P. E., Bos, S. D., Nelissen, R. G. & Meulenbelt, I. 2015. Transcriptional Associations of Osteoarthritis-Mediated Loss of Epigenetic Control in Articular Cartilage. *Arthritis Rheumatol,* 67**,** 2108-16.

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P. D., Dean, A. & Blobel, G. A. 2012. Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell,* 149**,** 1233-44.

Di Croce, L. & Helin, K. 2013. Transcriptional regulation by Polycomb group proteins. *Nat Struct Mol Biol,* 20**,** 1147-55.

Diao, Y., Li, B., Meng, Z., Jung, I., Lee, A. Y., Dixon, J., Maliskova, L., Guan, K. L., Shen, Y. & Ren, B. 2016. A new class of temporarily phenotypic enhancers identified by CRISPR/Cas9-mediated genetic screening. *Genome Res,* 26**,** 397-405.

Dipaola, C. P., Farmer, J. C., Manova, K. & Niswander, L. A. 2005. Molecular signaling in intervertebral disk development. *J Orthop Res,* 23**,** 1112-9.

Dirckx, N., Van Hul, M. & Maes, C. 2013. Osteoblast recruitment to sites of bone formation in skeletal development, homeostasis, and regeneration. *Birth Defects Res C Embryo Today,* 99**,** 170-91.

Doege, K., Hall, L. B., Mckinnon, W., Chen, L., Stephens, D. T. & Garrison, K. 2002. A remote upstream element regulates tissue-specific expression of the rat aggrecan gene. *J Biol Chem,* 277**,** 13989-97.

Doege, K. J., Garrison, K., Coulter, S. N. & Yamada, Y. 1994. The structure of the rat aggrecan gene and preliminary characterization of its promoter. *J Biol Chem,* 269**,** 29232-40.

Doege, K. J., Sasaki, M., Kimura, T. & Yamada, Y. 1991. Complete coding sequence and deduced primary structure of the human cartilage large aggregating proteoglycan, aggrecan. Human-specific repeats, and additional alternatively spliced forms. *J Biol Chem,* 266**,** 894-902.

Dolgova, E. V., Efremov, Y. R., Taranov, O. S., Potter, E. A., Nikolin, V. P., Popova, N. A., Omigov, V. V., Chernykh, E. R., Proskurina, A. S. & Bogachev, S. S. 2015. Comparative analysis of pathologic processes developing in mice housed in SPF

vs non-SPF conditions and treated with cyclophosphamide and dsDNA preparation. *Pathology - Research and Practice,* 211**,** 754-758.

Dominguez, A. A., Lim, W. A. & Qi, L. S. 2016. Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. *Nat Rev Mol Cell Biol,* 17**,** 5-15.

Domowicz, M., Krueger, R. C., Li, H., Mangoura, D., Vertel, B. M. & Schwartz, N. B. 1996. The nanomelic mutation in the aggrecan gene is expressed in chick chondrocytes and neurons. *Int J Dev Neurosci,* 14**,** 191-201.

Domowicz, M. S., Cortes, M., Henry, J. G. & Schwartz, N. B. 2009. Aggrecan modulation of growth plate morphogenesis. *Dev Biol,* 329**,** 242-57.

Domowicz, M. S., Sanders, T. A., Ragsdale, C. W. & Schwartz, N. B. 2008. Aggrecan is expressed by embryonic brain glia and regulates astrocyte development. *Dev Biol,* 315**,** 114-24.

Dow, L. E., Fisher, J., O'rourke, K. P., Muley, A., Kastenhuber, E. R., Livshits, G., Tschaharganeh, D. F., Socci, N. D. & Lowe, S. W. 2015. Inducible in vivo genome editing with CRISPR-Cas9. *Nat Biotechnol,* 33**,** 390-4.

Duan, X., Bradbury, S. R., Olsen, B. R. & Berendsen, A. D. 2016. VEGF stimulates intramembranous bone formation during craniofacial skeletal development. *Matrix Biol,* 52-54**,** 127-40.

Duan, X., Murata, Y., Liu, Y., Nicolae, C., Olsen, B. R. & Berendsen, A. D. 2015. Vegfa regulates perichondrial vascularity and osteoblast differentiation in bone development. *Development*.

Durigova, M., Nagase, H., Mort, J. S. & Roughley, P. J. 2011. MMPs are less efficient than ADAMTS5 in cleaving aggrecan core protein. *Matrix Biol,* 30**,** 145-53.

Dy, P., Penzo-Mendez, A., Wang, H., Pedraza, C., Macklin, W. & Lefebvre, V. 2008. The three SoxC proteins-Sox4, Sox11 and Sox12-exhibit overlapping expression patterns and molecular properties. *Nucleic Acids Res,* 36**,** 3101 - 3117.

Dy, P., Smits, P., Silvester, A., Penzo-Mendez, A., Dumitriu, B., Han, Y., De La Motte, C. A., Kingsley, D. M. & Lefebvre, V. 2010. Synovial joint morphogenesis requires

the chondrogenic action of Sox5 and Sox6 in growth plate and articular cartilage. *Dev Biol,* 341**,** 346-59.

Dy, P., Wang, W., Bhattaram, P., Wang, Q., Wang, L., Ballock, R. T. & Lefebvre, V. 2012. Sox9 directs hypertrophic maturation and blocks osteoblast differentiation of growth plate chondrocytes. *Dev Cell,* 22**,** 597-609.

Elgin, S. C. & Reuter, G. 2013. Position-effect variegation, heterochromatin formation, and gene silencing in Drosophila. *Cold Spring Harb Perspect Biol,* 5**,** a017780.

Ernst, J., Kheradpour, P., Mikkelsen, T. S., Shoresh, N., Ward, L. D., Epstein, C. B., Zhang, X., Wang, L., Issner, R., Coyne, M., Ku, M., Durham, T., Kellis, M. & Bernstein, B. E. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature,* 473**,** 43-9.

Esvelt, K. M., Mali, P., Braff, J. L., Moosburner, M., Yaung, S. J. & Church, G. M. 2013. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat Methods,* 10**,** 1116-21.

Eun, B., Sampley, M. L., Good, A. L., Gebert, C. M. & Pfeifer, K. 2013. Promoter cross-talk via a shared enhancer explains paternally biased expression of Nctc1 at the Igf2/H19/Nctc1 imprinted locus. *Nucleic Acids Res,* 41**,** 817-26.

Fang, M., Alfieri, C. M., Hulin, A., Conway, S. J. & Yutzey, K. E. 2014. Loss of beta-Catenin Promotes Chondrogenic Differentiation of Aortic Valve Interstitial Cells. *Arterioscler Thromb Vasc Biol,* 34**,** 2601-8.

Feil, R., Wagner, J., Metzger, D. & Chambon, P. 1997. Regulation of Cre Recombinase Activity by Mutated Estrogen Receptor Ligand-Binding Domains. *Biochemical and Biophysical Research Communications,* 237**,** 752-757.

Finger, F., Schorle, C., Zien, A., Gebhard, P., Goldring, M. B. & Aigner, T. 2003. Molecular phenotyping of human chondrocyte cell lines T/C-28a2, T/C-28a4, and C-28/I2. *Arthritis Rheum,* 48**,** 3395-403.

Fosang, A. J., Golub, S. B., East, C. J. & Rogerson, F. M. 2013. Abundant LacZ activity in the absence of Cre expression in the normal and inflamed synovium of adult Col2a1-Cre; ROSA26RLacZ reporter mice. *Osteoarthritis Cartilage,* 21**,** 401-4.

Franke, M., Ibrahim, D. M., Andrey, G., Schwarzer, W., Heinrich, V., Schöpflin, R., Kraft, K., Kempfer, R., Jerković, I., Chan, W.-L., Spielmann, M., Timmermann, B., Wittler, L., Kurth, I., Cambiaso, P., Zuffardi, O., Houge, G., Lambie, L., Brancati, F., Pombo, A., Vingron, M., Spitz, F. & Mundlos, S. 2016. Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature,* 538**,** 265-269.

Frankel, N., Davis, G. K., Vargas, D., Wang, S., Payre, F. & Stern, D. L. 2010. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature,* 466**,** 490-3.

Frischknecht, R. & Seidenbecher, C. I. 2012. Brevican: A key proteoglycan in the perisynaptic extracellular matrix of the brain. *The International Journal of Biochemistry & Cell Biology,* 44**,** 1051-1054.

Fujimaki, R., Hayashi, K., Watanabe, N., Yamada, T., Toyama, Y., Tezuka, K.-I. & Hozumi, N. 2005. Expression of Cre recombinase in the mouse developing chondrocytes driven by the mouse α2(XI) collagen promoter. *Journal of Bone and Mineral Metabolism,* 23**,** 270-273.

Fukada, K., Shibata, S., Suzuki, S., Ohya, K. & Kuroda, T. 1999. In situ hybridisation study of type I, II, X collagens and aggrecan mRNAs in the developing condylar cartilage of fetal mouse mandible. *Journal of Anatomy,* 195**,** 321-329.

Fukaya, T., Lim, B. & Levine, M. 2016. Enhancer Control of Transcriptional Bursting. *Cell,* 166**,** 358-368.

Galupa, R. & Heard, E. 2015. X-chromosome inactivation: new insights into cis and trans regulation. *Current Opinion in Genetics & Development,* 31**,** 57-66.

Gao, F., Shen, X. Z., Jiang, F., Wu, Y. & Han, C. 2016. DNA-guided genome editing using the Natronobacterium gregoryi Argonaute. *Nat Biotechnol*.

Gao, Y., Liu, S., Huang, J., Guo, W., Chen, J., Zhang, L., Zhao, B., Peng, J., Wang, A., Wang, Y., Xu, W., Lu, S., Yuan, M. & Guo, Q. 2014. The ECM-cell interaction of cartilage extracellular matrix on chondrocytes. *Biomed Res Int,* 2014**,** 648459.

Garside, V. C., Cullum, R., Alder, O., Lu, D. Y., Vander Werff, R., Bilenky, M., Zhao, Y., Jones, S. J., Marra, M. A., Underhill, T. M. & Hoodless, P. A. 2015. SOX9 modulates the expression of key transcription factors required for heart valve development. *Development,* 142**,** 4340-50.

Gaszner, M. & Felsenfeld, G. 2006. Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet,* 7**,** 703-13.

Gebauer, M., Saas, J., Sohler, F., Haag, J., Söder, S., Pieper, M., Bartnik, E., Beninga, J., Zimmer, R. & Aigner, T. 2005. Comparison of the chondrosarcoma cell line SW1353 with primary human adult articular chondrocytes with regard to their gene expression profile and reactivity to IL-1β. *Osteoarthritis and Cartilage,* 13**,** 697-708.

Gebhard, S., Pöschl, E., Riemer, S., Bauer, E., Hattori, T., Eberspaecher, H., Zhang, Z., Lefebvre, V., De Crombrugghe, B. & Von Der Mark, K. 2004. A highly conserved enhancer in mammalian type X collagen genes drives high levels of tissue-specific expression in hypertrophic cartilage in vitro and in vivo. *Matrix Biology,* 23**,** 309-322.

Gee, F., Clubbs, C. F., Raine, E. V., Reynard, L. N. & Loughlin, J. 2014. Allelic expression analysis of the osteoarthritis susceptibility locus that maps to chromosome 3p21 reveals cis-acting eQTLs at GNL3 and SPCS1. *BMC Med Genet,* 15**,** 53.

Gershenzon, N. I. & Ioshikhes, I. P. 2005. Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. *Bioinformatics,* 21**,** 1295-300.

Ghirlando, R., Giles, K., Gowher, H., Xiao, T., Xu, Z., Yao, H. & Felsenfeld, G. 2012. Chromatin domains, insulators, and the regulation of gene expression. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms,* 1819**,** 644-651.

Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., Stern-Ginossar, N., Brandman, O., Whitehead, E. H., Doudna, J. A., Lim, W. A., Weissman, J. S. & Qi, L. S. 2013. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell,* 154**,** 442-51.

Gleghorn, L., Ramesar, R., Beighton, P. & Wallis, G. 2005. A mutation in the variable repeat region of the aggrecan gene (AGC1) causes a form of spondyloepiphyseal dysplasia associated with severe, premature osteoarthritis. *Am J Hum Genet,* 77**,** 484-90.

Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A. & Kornberg, R. D. 2001. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. *Science,* 292**,** 1876-82.

Goldring, M. B. 2012. Chondrogenesis, chondrocyte differentiation, and articular cartilage metabolism in health and osteoarthritis. *Therapeutic Advances in Musculoskeletal Disease,* 4**,** 269-285.

Goldring, M. B., Birkhead, J. R., Suen, L. F., Yamin, R., Mizuno, S., Glowacki, J., Arbiser, J. L. & Apperley, J. F. 1994. Interleukin-1 beta-modulated gene expression in immortalized human chondrocytes. *J Clin Invest,* 94**,** 2307-16.

Goldring, M. B. & Otero, M. 2011. Inflammation in osteoarthritis. *Curr Opin Rheumatol,* 23**,** 471-8.

Goldring, M. B., Tsuchimochi, K. & Ijiri, K. 2006. The control of chondrogenesis. *J Cell Biochem,* 97**,** 33-44.

Goloborodko, A., Imakaev, M. V., Marko, J. F. & Mirny, L. 2016a. Compaction and segregation of sister chromatids via active loop extrusion. *eLife,* 5**,** e14864.

Goloborodko, A., Marko, John f. & Mirny, Leonid a. 2016b. Chromosome Compaction by Active Loop Extrusion. *Biophysical Journal,* 110**,** 2162-2168.

Gonzalez, F., Duboule, D. & Spitz, F. 2007. Transgenic analysis of Hoxd gene regulation during digit development. *Developmental Biology,* 306**,** 847-859.

Goodnough, L. H., Chang, A. T., Treloar, C., Yang, J., Scacheri, P. C. & Atit, R. P. 2012. Twist1 mediates repression of chondrogenesis by beta-catenin to promote cranial bone progenitor specification. *Development,* 139**,** 4428-38.

Goodnough, L. H., Dinuoscio, G. J., Ferguson, J. W., Williams, T., Lang, R. A. & Atit, R. P. 2014. Distinct requirements for cranial ectoderm and mesenchyme-derived wnts

in specification and differentiation of osteoblast and dermal progenitors. *PLoS Genet,* 10**,** e1004152.

Gu, J., Lu, Y., Li, F., Qiao, L., Wang, Q., Li, N., Borgia, J. A., Deng, Y., Lei, G. & Zheng, Q. 2014. Identification and characterization of the novel Col10a1 regulatory mechanism during chondrocyte hypertrophic differentiation. *Cell Death Dis,* 5**,** e1469.

Gualeni, B., Facchini, M., De Leonardis, F., Tenni, R., Cetta, G., Viola, M., Passi, A., Superti-Furga, A., Forlino, A. & Rossi, A. 2010. Defective proteoglycan sulfation of the growth plate zones causes reduced chondrocyte proliferation via an altered Indian hedgehog signalling. *Matrix Biol,* 29**,** 453-60.

Hall, M. A., Shundrovsky, A., Bai, L., Fulbright, R. M., Lis, J. T. & Wang, M. D. 2009. High-resolution dynamic mapping of histone-DNA interactions in a nucleosome. *Nat Struct Mol Biol,* 16**,** 124-9.

Han, Y. & Lefebvre, V. 2008. L-Sox5 and Sox6 drive expression of the aggrecan gene in cartilage by securing binding of Sox9 to a far-upstream enhancer. *Mol Cell Biol,* 28**,** 4999-5013.

Hattori, T., Muller, C., Gebhard, S., Bauer, E., Pausch, F., Schlund, B., Bosl, M. R., Hess, A., Surmann-Schmitt, C., Von Der Mark, H., De Crombrugghe, B. & Von Der Mark, K. 2010. SOX9 is a major negative regulator of cartilage vascularization, bone marrow formation and endochondral ossification. *Development,* 137**,** 901-11.

Hay, D., Hughes, J. R., Babbs, C., Davies, J. O. J., Graham, B. J., Hanssen, L. L. P., Kassouf, M. T., Oudelaar, A. M., Sharpe, J. A., Suciu, M. C., Telenius, J., Williams, R., Rode, C., Li, P.-S., Pennacchio, L. A., Sloane-Stanley, J. A., Ayyub, H., Butler, S., Sauka-Spengler, T., Gibbons, R. J., Smith, A. J. H., Wood, W. G. & Higgs, D. R. 2016. Genetic dissection of the α-globin super-enhancer in vivo. *Nature Genetics*.

Heffner, C. S., Herbert Pratt, C., Babiuk, R. P., Sharma, Y., Rockwood, S. F., Donahue, L. R., Eppig, J. T. & Murray, S. A. 2012. Supporting conditional mouse mutagenesis with a comprehensive cre characterization resource. *Nat Commun,* 3**,** 1218.

Heinegard, D. 2009. Proteoglycans and more--from molecules to biology. *Int J Exp Pathol,* 90**,** 575-86.

Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., Ye, Z., Lee, L. K., Stuart, R. K., Ching, C. W., Ching, K. A., Antosiewicz-Bourget, J. E., Liu, H., Zhang, X., Green, R. D., Lobanenkov, V. V., Stewart, R., Thomson, J. A., Crawford, G. E. & Kellis, M. 2009a. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature,* 459**,** 108-112.

Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., Ye, Z., Lee, L. K., Stuart, R. K., Ching, C. W., Ching, K. A., Antosiewicz-Bourget, J. E., Liu, H., Zhang, X., Green, R. D., Lobanenkov, V. V., Stewart, R., Thomson, J. A., Crawford, G. E., Kellis, M. & Ren, B. 2009b. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature,* 459**,** 108-12.

Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., Wang, W., Weng, Z., Green, R. D., Crawford, G. E. & Ren, B. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet,* 39**,** 311-8.

Henikoff, S. & Shilatifard, A. 2011. Histone modification: cause or cog? *Trends Genet,* 27**,** 389-96.

Henry, S. P., Jang, C.-W., Deng, J. M., Zhang, Z., Behringer, R. R. & De Crombrugghe, B. 2009. Generation of aggrecan-CreERT2 knockin mice for inducible Cre activity in adult cartilage. *genesis,* 47**,** 805-814.

Hering, T. M., Kollar, J. & Huynh, T. D. 1997. Complete Coding Sequence of Bovine Aggrecan: Comparative Structural Analysis. *Archives of Biochemistry and Biophysics,* 345**,** 259-270.

Hernandez-Garcia, C. M. & Finer, J. J. 2014. Identification and validation of promoters and cis-acting regulatory elements. *Plant Science,* 217–218**,** 109-119.

Herrmann, J., Gressner, A. M. & Weiskirchen, R. 2007. Immortal hepatic stellate cell lines: useful tools to study hepatic stellate cell biology and function? *J Cell Mol Med,* 11**,** 704-22.

Hilton, M. J., Tu, X., Cook, J., Hu, H. & Long, F. 2005. Ihh controls cartilage development by antagonizing Gli3, but requires additional effectors to regulate osteoblast and vascular development. *Development,* 132**,** 4339-51.

Hilton, M. J., Tu, X. & Long, F. 2007. Tamoxifen-inducible gene deletion reveals a distinct cell type associated with trabecular bone, and direct regulation of PTHrP expression and chondrocyte morphology by Ihh in growth region cartilage. *Dev Biol,* 308**,** 93-105.

Hilton, M. J., Tu, X., Wu, X., Bai, S., Zhao, H., Kobayashi, T., Kronenberg, H. M., Teitelbaum, S. L., Ross, F. P., Kopan, R. & Long, F. 2008. Notch signaling maintains bone marrow mesenchymal progenitors by suppressing osteoblast differentiation. *Nat Med,* 14**,** 306-314.

Hiraoka, S., Furuichi, T., Nishimura, G., Shibata, S., Yanagishita, M., Rimoin, D. L., Superti-Furga, A., Nikkels, P. G., Ogawa, M., Katsuyama, K., Toyoda, H., Kinoshita-Toyoda, A., Ishida, N., Isono, K., Sanai, Y., Cohn, D. H., Koseki, H. & Ikegawa, S. 2007. Nucleotide-sugar transporter SLC35D1 is critical to chondroitin sulfate synthesis in cartilage and skeletal development in mouse and human. *Nat Med,* 13**,** 1363-7.

Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-Andre, V., Sigova, A. A., Hoke, H. A. & Young, R. A. 2013. Super-enhancers in the control of cell identity and disease. *Cell,* 155**,** 934-47.

Hnisz, D., Schuijers, J., Lin, C. Y., Weintraub, A. S., Abraham, B. J., Lee, T. I., Bradner, J. E. & Young, R. A. 2015. Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. *Mol Cell,* 58**,** 362-70.

Hobert, O. 2010. Gene regulation: enhancers stepping out of the shadow. *Curr Biol,* 20**,** R697-9.

Hogan, B., Beddington, R., Costantini, F. & Lacy, E. 1994. *Manipulating the Mouse Embryo, A laboratory manual*, Cold Sprin Harbor Laboatory Press.

Hong, J.-W., Hendrix, D. A. & Levine, M. S. 2008. Shadow Enhancers as a Source of Evolutionary Novelty. *Science,* 321**,** 1314-1314.

Hong, S., Derfoul, A., Pereira-Mouries, L. & Hall, D. J. 2009. A novel domain in histone deacetylase 1 and 2 mediates repression of cartilage-specific genes in human chondrocytes. *Faseb j,* 23**,** 3539-52.

Hu, G., Codina, M. & Fisher, S. 2012. Multiple enhancers associated with ACAN suggest highly redundant transcriptional regulation in cartilage. *Matrix Biol,* 31**,** 328-37.

Hu, G., Cui, K., Northrup, D., Liu, C., Wang, C., Tang, Q., Ge, K., Levens, D., Crane-Robinson, C. & Zhao, K. 2013. H2A.Z facilitates access of active and repressive complexes to chromatin in embryonic stem cell self-renewal and differentiation. *Cell Stem Cell,* 12**,** 180-92.

Huang, J., Liu, X., Li, D., Shao, Z., Cao, H., Zhang, Y., Trompouki, E., Bowman, T. V., Zon, L. I., Yuan, G. C., Orkin, S. H. & Xu, J. 2016. Dynamic Control of Enhancer Repertoires Drives Lineage and Stage-Specific Transcription during Hematopoiesis. *Dev Cell,* 36**,** 9-23.

Huang, X., Wang, D., Weiss, D. R., Bushnell, D. A., Kornberg, R. D. & Levitt, M. 2010. RNA polymerase II trigger loop residues stabilize and position the incoming nucleotide triphosphate in transcription. *Proc Natl Acad Sci U S A,* 107**,** 15745-50.

Hwang, W. Y., Fu, Y., Reyon, D., Maeder, M. L., Tsai, S. Q., Sander, J. D., Peterson, R. T., Yeh, J. R. & Joung, J. K. 2013. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol,* 31**,** 227-9.

Ikeda, T., Kamekura, S., Mabuchi, A., Kou, I., Seki, S., Takato, T., Nakamura, K., Kawaguchi, H., Ikegawa, S. & Chung, U. 2004. The combination of SOX5, SOX6, and SOX9 (the SOX trio) provides signals sufficient for induction of permanent cartilage. *Arthritis Rheum,* 50**,** 3561 - 3573.

Ikeda, Y., Ito, K., Izumi, Y. & Shinomura, T. 2014. A candidate enhancer element responsible for high-level expression of the aggrecan gene in chondrocytes. *J Biochem*.

Indra, A. K., Warot, X., Brocard, J., Bornert, J. M., Xiao, J. H., Chambon, P. & Metzger, D. 1999. Temporally-controlled site-specific mutagenesis in the basal layer of the

epidermis: comparison of the recombinase activity of the tamoxifen-inducible Cre-ER(T) and Cre-ER(T2) recombinases. *Nucleic Acids Res,* 27**,** 4324-7.

Ismail, H. M., Miotla-Zarebska, J., Troeberg, L., Tang, X., Stott, B., Yamamoto, K., Nagase, H., Fosang, A. J., Vincent, T. L. & Saklatvala, J. 2015a. JNK2 controls aggrecan degradation in murine articular cartilage and the development of experimental osteoarthritis. *Arthritis & Rheumatology***,** n/a-n/a.

Ismail, H. M., Yamamoto, K., Vincent, T. L., Nagase, H., Troeberg, L. & Saklatvala, J. 2015b. Interleukin-1 Acts via the JNK-2 Signaling Pathway to Induce Aggrecan Degradation by Human Chondrocytes. *Arthritis Rheumatol,* 67**,** 1826-36.

Jenkins, E., Moss, J. B., Pace, J. M. & Bridgewater, L. C. 2005. The new collagen gene COL27A1 contains SOX9-responsive enhancer elements. *Matrix Biol,* 24**,** 177-84.

Jeong, Y. & Epstein, D. J. 2003. Distinct regulators of Shh transcription in the floor plate and notochord indicate separate origins for these tissues in the mouse node. *Development,* 130**,** 3891-902.

Jiang, Z., Von Den Hoff, J. W., Torensma, R., Meng, L. & Bian, Z. 2014. Wnt16 is involved in intramembranous ossification and suppresses osteoblast differentiation through the Wnt/beta-catenin pathway. *J Cell Physiol,* 229**,** 384-92.

Jin, F., Li, Y., Dixon, J. R., Selvaraj, S., Ye, Z., Lee, A. Y., Yen, C.-A., Schmitt, A. D., Espinoza, C. A. & Ren, B. 2013. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature,* 503**,** 290-294.

Johnson, A., Wu, R., Peetz, M., Gygi, S. P. & Moazed, D. 2013. Heterochromatic gene silencing by activator interference and a transcription elongation barrier. *J Biol Chem,* 288**,** 28771-82.

Johnson, Z. I., Shapiro, I. M. & Risbud, M. V. 2014. Extracellular Osmolarity Regulates Matrix Homeostasis in the Intervertebral Disc and Articular Cartilage: Evolving role of TonEBP. *Matrix Biol*.

Juven-Gershon, T., Hsu, J. Y., Theisen, J. W. & Kadonaga, J. T. 2008. The RNA polymerase II core promoter - the gateway to transcription. *Curr Opin Cell Biol,* 20**,** 253-9.

Kadonaga, J. T. 2012. Perspectives on the RNA polymerase II core promoter. *Wiley Interdiscip Rev Dev Biol,* 1**,** 40-51.

Kagey, M. H., Newman, J. J., Bilodeau, S., Zhan, Y., Orlando, D. A., Van Berkum, N. L., Ebmeier, C. C., Goossens, J., Rahl, P. B., Levine, S. S., Taatjes, D. J., Dekker, J. & Young, R. A. 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature,* 467**,** 430-5.

Kalocsay, M., Hiller, N. J. & Jentsch, S. 2009. Chromosome-wide Rad51 spreading and SUMO-H2A.Z-dependent chromosome fixation in response to a persistent DNA double-strand break. *Mol Cell,* 33**,** 335-43.

Karlic, R., Chung, H. R., Lasserre, J., Vlahovicek, K. & Vingron, M. 2010. Histone modification levels are predictive for gene expression. *Proc Natl Acad Sci U S A,* 107**,** 2926-31.

Kelleher, R. J., 3rd, Flanagan, P. M. & Kornberg, R. D. 1990. A novel mediator between activator proteins and the RNA polymerase II transcription apparatus. *Cell,* 61**,** 1209-15.

Kerkhof, H. J., Lories, R. J., Meulenbelt, I., Jonsdottir, I., Valdes, A. M., Arp, P., Ingvarsson, T., Jhamai, M., Jonsson, H., Stolk, L., Thorleifsson, G., Zhai, G., Zhang, F., Zhu, Y., Van Der Breggen, R., Carr, A., Doherty, M., Doherty, S., Felson, D. T., Gonzalez, A., Halldorsson, B. V., Hart, D. J., Hauksson, V. B., Hofman, A., Ioannidis, J. P., Kloppenburg, M., Lane, N. E., Loughlin, J., Luyten, F. P. & Nevitt, M. C. 2010. A genome-wide association study identifies an osteoarthritis susceptibility locus on chromosome 7q22. *Arthritis Rheum,* 62.

Kiani, C., Chen, L., Wu, Y. J., Yee, A. J. & Yang, B. B. 2002. Structure and function of aggrecan. *Cell Res,* 12**,** 19-32.

Kim, K. W., Lim, T. H., Kim, J. G., Jeong, S. T., Masuda, K. & An, H. S. 2003. The origin of chondrocytes in the nucleus pulposus and histologic findings associated with the transition of a notochordal nucleus pulposus to a fibrocartilaginous nucleus pulposus in intact rabbit intervertebral discs. *Spine (Phila Pa 1976),* 28**,** 982-90.

Kim, Y. W., Lee, S., Yun, J. & Kim, A. 2015. Chromatin looping and eRNA transcription precede the transcriptional activation of gene in the beta-globin locus. *Biosci Rep,* 35.

Kletsas, D., Mccann, M. R., Patel, P., Frimpong, A., Xiao, Y., Siqueira, W. L. & Séguin, C. A. 2015. Proteomic Signature of the Murine Intervertebral Disc. *Plos One,* 10**,** e0117807.

Kmita, M., Tarchini, B., Zakany, J., Logan, M., Tabin, C. J. & Duboule, D. 2005. Early developmental arrest of mammalian limbs lacking HoxA/HoxD gene function. *Nature,* 435**,** 1113-1116.

Kohn, A., Dong, Y., Mirando, A. J., Jesse, A. M., Honjo, T., Zuscik, M. J., O'keefe, R. J. & Hilton, M. J. 2012. Cartilage-specific RBPjkappa-dependent and -independent Notch signals regulate cartilage and bone development. *Development,* 139**,** 1198-212.

Kohn, A., Rutkowski, T. P., Liu, Z., Mirando, A. J., Zuscik, M. J., O'keefe, R. J. & Hilton, M. J. 2015. Notch signaling controls chondrocyte hypertrophy via indirect regulation of Sox9. *Bone Res,* 3**,** 15021.

Kolpakova-Hart, E., Nicolae, C., Zhou, J. & Olsen, B. R. 2008. Col2-Cre recombinase is co-expressed with endogenous type II collagen in embryonic renal epithelium and drives development of polycystic kidney disease following inactivation of ciliary genes. *Matrix biology : journal of the International Society for Matrix Biology,* 27**,** 505-512.

Konermann, S., Brigham, M. D., Trevino, A. E., Joung, J., Abudayyeh, O. O., Barcena, C., Hsu, P. D., Habib, N., Gootenberg, J. S., Nishimasu, H., Nureki, O. & Zhang, F. 2015. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature,* 517**,** 583-8.

Korkmaz, G., Lopes, R., Ugalde, A. P., Nevedomskaya, E., Han, R., Myacheva, K., Zwart, W., Elkon, R. & Agami, R. 2016. Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat Biotechnol,* 34**,** 192-8.

Kornberg, R. D. 2005. Mediator and the mechanism of transcriptional activation. *Trends Biochem Sci,* 30**,** 235-9.

Kornberg, R. D. 2007. The molecular basis of eukaryotic transcription. *Proc Natl Acad Sci U S A,* 104**,** 12955-61.

Kothary, R., Clapoff, S., Darling, S., Perry, M. D., Moran, L. A. & Rossant, J. 1989. Inducible expression of an hsp68-lacZ hybrid gene in transgenic mice. *Development,* 105**,** 707-14.

Kouzarides, T. 2007. Chromatin modifications and their function. *Cell,* 128**,** 693-705.

Kowalczyk, M. S., Hughes, J. R., Garrick, D., Lynch, M. D., Sharpe, J. A., Sloane-Stanley, J. A., Mcgowan, S. J., De Gobbi, M., Hosseini, M., Vernimmen, D., Brown, J. M., Gray, N. E., Collavin, L., Gibbons, R. J., Flint, J., Taylor, S., Buckle, V. J., Milne, T. A., Wood, W. G. & Higgs, D. R. 2012. Intragenic enhancers act as alternative promoters. *Mol Cell,* 45**,** 447-58.

Koyama, E., Ochiai, T., Rountree, R. B., Kingsley, D. M., Enomoto-Iwamoto, M., Iwamoto, M. & Pacifici, M. 2007. Synovial joint formation during mouse limb skeletogenesis: roles of Indian hedgehog signaling. *Ann N Y Acad Sci,* 1116**,** 100-12.

Koyama, E., Shibukawa, Y., Nagayama, M., Sugito, H., Young, B., Yuasa, T., Okabe, T., Ochiai, T., Kamiya, N., Rountree, R. B., Kingsley, D. M., Iwamoto, M., Enomoto-Iwamoto, M. & Pacifici, M. 2008. A distinct cohort of progenitor cells participates in synovial joint and articular cartilage formation during mouse limb skeletogenesis. *Dev Biol,* 316**,** 62-73.

Kozhemyakina, E., Lassar, A. B. & Zelzer, E. 2015a. A pathway to bone: signaling molecules and transcription factors involved in chondrocyte development and maturation. *Development,* 142**,** 817-31.

Kozhemyakina, E., Zhang, M., Ionescu, A., Ayturk, U. M., Ono, N., Kobayashi, A., Kronenberg, H., Warman, M. L. & Lassar, A. B. 2015b. Identification of a Prg4-expressing articular cartilage progenitor cell population in mice. *Arthritis Rheumatol,* 67**,** 1261-73.

Koziel, L., Wuelling, M., Schneider, S. & Vortkamp, A. 2005. Gli3 acts as a repressor downstream of Ihh in regulating two distinct steps of chondrocyte differentiation. *Development,* 132**,** 5249-60.

Kresse, H. & Schonherr, E. 2001. Proteoglycans of the extracellular matrix and growth control. *J Cell Physiol,* 189**,** 266-74.

Krivega, I. & Dean, A. 2012. Enhancer and promoter interactions-long distance calls. *Curr Opin Genet Dev,* 22**,** 79-85.

Krivega, I. & Dean, A. 2016. Chromatin looping as a target for altering erythroid gene expression. *Ann N Y Acad Sci,* 1368**,** 31-9.

Krueger, R. C., Jr., Kurima, K. & Schwartz, N. B. 1999. Completion of the mouse aggrecan gene structure and identification of the defect in the cmd-Bc mouse as a near complete deletion of the murine aggrecan gene. *Mamm Genome,* 10**,** 1119-25.

Krull, N. B. & Gressner, A. M. 1992. Differential expression of keratan sulphate proteoglycans fibromodulin, lumican and aggrecan in normal and fibrotic rat liver. *FEBS Lett,* 312**,** 47-52.

Kvon, E. Z. 2015. Using transgenic reporter assays to functionally characterize enhancers in animals. *Genomics,* 106**,** 185-192.

Lafont, J. E., Talma, S., Hopfgarten, C. & Murphy, C. L. 2008. Hypoxia promotes the differentiated human articular chondrocyte phenotype through SOX9-dependent and -independent pathways. *J Biol Chem,* 283**,** 4778-86.

Lafont, J. E., Talma, S. & Murphy, C. L. 2007. Hypoxia-inducible factor 2alpha is essential for hypoxic induction of the human articular chondrocyte phenotype. *Arthritis Rheum,* 56**,** 3297-306.

Lantermann, A. B., Straub, T., Stralfors, A., Yuan, G. C., Ekwall, K. & Korber, P. 2010. Schizosaccharomyces pombe genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of Saccharomyces cerevisiae. *Nat Struct Mol Biol,* 17**,** 251-7.

Larson, M. H., Zhou, J., Kaplan, C. D., Palangat, M., Kornberg, R. D., Landick, R. & Block, S. M. 2012. Trigger loop dynamics mediate the balance between the

transcriptional fidelity and speed of RNA polymerase II. *Proceedings of the National Academy of Sciences,* 109**,** 6555-6560.

Lauer, A. M., May, B. J., Hao, Z. J. & Watson, J. 2009. Sound levels in modern rodent housing rooms are an uncontrolled environmental variable with fluctuations mainly due to human activities. *Lab animal,* 38**,** 154-160.

Lawrence, M., Daujat, S. & Schneider, R. 2016. Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends Genet,* 32**,** 42-56.

Lee, G. R., Fields, P. E., Griffin, T. J. & Flavell, R. A. 2003. Regulation of the Th2 cytokine locus by a locus control region. *Immunity,* 19**,** 145-53.

Lee, H. S., Park, J. H., Kim, S. J., Kwon, S. J. & Kwon, J. 2010. A cooperative activation loop among SWI/SNF, γ-H2AX and H3 acetylation for DNA double-strand break repair. *The EMBO Journal,* 29**,** 1434-1445.

Lee, N. V., Rodriguez-Manzaneque, J. C., Thai, S. N., Twal, W. O., Luque, A., Lyons, K. M., Argraves, W. S. & Iruela-Arispe, M. L. 2005. Fibulin-1 acts as a cofactor for the matrix metalloprotease ADAMTS-1. *J Biol Chem,* 280**,** 34796-804.

Lefebvre, V., Huang, W., Harley, V. R., Goodfellow, P. N. & De Crombrugghe, B. 1997. SOX9 is a potent activator of the chondrocyte-specific enhancer of the pro alpha1(II) collagen gene. *Mol Cell Biol,* 17**,** 2336-46.

Lefebvre, V., Li, P. & De Crombrugghe, B. 1998. A new long form of Sox5 (L-Sox5), Sox6 and Sox9 are coexpressed in chondrogenesis and cooperatively activate the type II collagen gene. *EMBO J,* 17**,** 5718-33.

Lefebvre, V., Zhou, G., Mukhopadhyay, K., Smith, C. N., Zhang, Z., Eberspaecher, H., Zhou, X., Sinha, S., Maity, S. N. & De Crombrugghe, B. 1996. An 18-base-pair sequence in the mouse proalpha1(II) collagen gene is sufficient for expression in cartilage and binds nuclear proteins that are selectively expressed in chondrocytes. *Mol Cell Biol,* 16**,** 4512-23.

Lelli, K. M., Slattery, M. & Mann, R. S. 2012. Disentangling the many layers of eukaryotic transcriptional regulation. *Annu Rev Genet,* 46**,** 43-68.

Lenton, K., James, A. W., Manu, A., Brugmann, S. A., Birker, D., Nelson, E. R., Leucht, P., Helms, J. A. & Longaker, M. T. 2011. Indian hedgehog positively regulates calvarial ossification and modulates bone morphogenetic protein signaling. *Genesis,* 49**,** 784-96.

Leung, K. K., Ng, L. J., Ho, K. K., Tam, P. P. & Cheah, K. S. 1998. Different cis-regulatory DNA elements mediate developmental stage- and tissue-specific expression of the human COL2A1 gene in transgenic mice. *J Cell Biol,* 141**,** 1291-300.

Leung, V. Y., Gao, B., Leung, K. K., Melhado, I. G., Wynn, S. L., Au, T. Y., Dung, N. W., Lau, J. Y., Mak, A. C., Chan, D. & Cheah, K. S. 2011. SOX9 governs differentiation stage-specific gene expression in growth plate chondrocytes via direct concomitant transactivation and repression. *PLoS Genet,* 7**,** e1002356.

Levine, M., Cattoglio, C. & Tjian, R. 2014. Looping back to leap forward: transcription enters a new era. *Cell,* 157**,** 13-25.

Lewis, B. A., Kim, T. K. & Orkin, S. H. 2000. A downstream element in the human beta-globin promoter: evidence of extended sequence-specific transcription factor IID contacts. *Proc Natl Acad Sci U S A,* 97**,** 7172-7.

Li, B., Carey, M. & Workman, J. L. 2007. The role of chromatin during transcription. *Cell,* 128**,** 707-19.

Li, H. & Schwartz, N. 1995. Gene structure of chick cartilage chondroitin sulfate proteoglycan (aggrecan) core protein. *Journal of Molecular Evolution,* 41**,** 878-885.

Li, H., Schwartz, N. B. & Vertel, B. M. 1993. cDNA cloning of chick cartilage chondroitin sulfate (aggrecan) core protein and identification of a stop codon in the aggrecan gene associated with the chondrodystrophy, nanomelia. *J Biol Chem,* 268**,** 23504-11.

Li, L. M. & Arnosti, D. N. 2011. Long- and Short-Range Transcriptional Repressors Induce Distinct Chromatin States on Repressed Genes. *Current Biology,* 21**,** 406-412.

Li, Q., Peterson, K. R., Fang, X. & Stamatoyannopoulos, G. 2002. Locus control regions. *Blood,* 100**,** 3077-86.

Li, T., Longobardi, L., Myers, T. J., Temple, J. D., Chandler, R. L., Ozkan, H., Contaldo, C. & Spagnoli, A. 2013. Joint TGF-beta type II receptor-expressing cells: ontogeny and characterization as joint progenitors. *Stem Cells Dev,* 22**,** 1342-59.

Li, Y., Rivera, C. M., Ishii, H., Jin, F., Selvaraj, S., Lee, A. Y., Dixon, J. R. & Ren, B. 2014. CRISPR reveals a distal super-enhancer required for Sox2 expression in mouse embryonic stem cells. *PLoS One,* 9**,** e114485.

Liao, J., Hu, N., Zhou, N., Lin, L., Zhao, C., Yi, S., Fan, T., Bao, W., Liang, X., Chen, H., Xu, W., Chen, C., Cheng, Q., Zeng, Y., Si, W., Yang, Z. & Huang, W. 2014. Sox9 Potentiates BMP2-Induced Chondrogenic Differentiation and Inhibits BMP2-Induced Osteogenic Differentiation. *PLoS One,* 9**,** e89025.

Lim, C. Y., Santoso, B., Boulay, T., Dong, E., Ohler, U. & Kadonaga, J. T. 2004. The MTE, a new core promoter element for transcription by RNA polymerase II. *Genes Dev,* 18**,** 1606-17.

Lin, S. S., Tzeng, B. H., Lee, K. R., Smith, R. J., Campbell, K. P. & Chen, C. C. 2014. Cav3.2 T-type calcium channel is required for the NFAT-dependent Sox9 expression in tracheal cartilage. *Proc Natl Acad Sci U S A,* 111**,** E1990-8.

Lincoln, J., Alfieri, C. M. & Yutzey, K. E. 2006a. BMP and FGF regulatory pathways control cell lineage diversification of heart valve precursor cells. *Developmental Biology,* 292**,** 290-302.

Lincoln, J., Lange, A. W. & Yutzey, K. E. 2006b. Hearts and bones: shared regulatory mechanisms in heart valve, cartilage, tendon, and bone development. *Dev Biol,* 294**,** 292-302.

Link, N., Kurtz, P., O'neal, M., Garcia-Hughes, G. & Abrams, J. M. 2013. A p53 enhancer region regulates target genes through chromatin conformations in cis and in trans. *Genes Dev,* 27**,** 2433-8.

Little, C. B., Meeker, C. T., Hembry, R. M., Sims, N. A., Lawlor, K. E., Golub, S. B., Last, K. & Fosang, A. J. 2005. Matrix Metalloproteinases Are Not Essential for Aggrecan Turnover during Normal Skeletal Growth and Development. *Molecular and Cellular Biology,* 25**,** 3388-3399.

Little, D. G., Ramachandran, M. & Schindeler, A. 2007. The anabolic and catabolic responses in bone repair. *Bone & Joint Journal,* 89-B**,** 425-433.

Liu, C. F. & Lefebvre, V. 2015. The transcription factors SOX9 and SOX5/SOX6 cooperate genome-wide through super-enhancers to drive chondrogenesis. *Nucleic Acids Res*.

Liu, X., Bushnell, D. A., Wang, D., Calero, G. & Kornberg, R. D. 2010. Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism. *Science,* 327**,** 206-9.

Liu, Z., Chen, J., Mirando, A. J., Wang, C., Zuscik, M. J., O'keefe, R. J. & Hilton, M. J. 2015. A dual role for NOTCH signaling in joint cartilage maintenance and osteoarthritis. *Sci Signal,* 8**,** ra71.

Liu, Z., Ren, Y., Mirando, A. J., Wang, C., Zuscik, M. J., O'keefe, R. J. & Hilton, M. J. 2016. Notch signaling in postnatal joint chondrocytes, but not subchondral osteoblasts, is required for articular cartilage and joint maintenance. *Osteoarthritis Cartilage,* 24**,** 740-51.

Logan, M., Martin, J. F., Nagy, A., Lobe, C., Olson, E. N. & Tabin, C. J. 2002. Expression of Cre Recombinase in the developing mouse limb bud driven by a Prxl enhancer. *Genesis,* 33**,** 77-80.

Lojda, Z. 1970. Indigogenic methods for glycosidases. I. An improved method for beta-D-glucosidase and its application to localization studies on intestinal and renal enzymes. *Histochemie,* 22**,** 347-61.

Long, F., Zhang, X. M., Karp, S., Yang, Y. & Mcmahon, A. P. 2001. Genetic manipulation of hedgehog signaling in the endochondral skeleton reveals a direct role in the regulation of chondrocyte proliferation. *Development,* 128**,** 5099-5108.

Lorch, Y., Davis, B. & Kornberg, R. D. 2005. Chromatin remodeling by DNA bending, not twisting. *Proc Natl Acad Sci U S A,* 102**,** 1329-32.

Lorch, Y., Maier-Davis, B. & Kornberg, R. D. 2010. Mechanism of chromatin remodeling. *Proc Natl Acad Sci U S A,* 107**,** 3458-62.

Lorda-Diez, C. I., Montero, J. A., Rodriguez-Leon, J., Garcia-Porrero, J. A. & Hurle, J. M. 2013. Expression and functional study of extracellular BMP antagonists during the morphogenesis of the digits and their associated connective tissues. *PLoS One,* 8**,** e60423.

Loven, J., Hoke, H. A., Lin, C. Y., Lau, A., Orlando, D. A., Vakoc, C. R., Bradner, J. E., Lee, T. I. & Young, R. A. 2013. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell,* 153**,** 320-34.

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. 1997. Crystal structure of the nucleosome core particle at 2.8[thinsp]A resolution. *Nature,* 389**,** 251-260.

Luo, W., Kuwada, T. S., Chandrasekaran, L., Zheng, J. & Tanzer, M. L. 1996. Divergent Secretory Behavior of the Opposite Ends of Aggrecan. *Journal of Biological Chemistry,* 271**,** 16447-16450.

Ma, E., Harrington, L. B., O'connell, M. R., Zhou, K. & Doudna, J. A. 2015a. Single-Stranded DNA Cleavage by Divergent CRISPR-Cas9 Enzymes. *Mol Cell,* 60**,** 398-407.

Ma, H., Naseri, A., Reyes-Gutierrez, P., Wolfe, S. A., Zhang, S. & Pederson, T. 2015b. Multicolor CRISPR labeling of chromosomal loci in human cells. *Proceedings of the National Academy of Sciences,* 112**,** 3002-3007.

Mackie, E. J., Ahmed, Y. A., Tatarczuch, L., Chen, K. S. & Mirams, M. 2008. Endochondral ossification: how cartilage is converted into bone in the developing skeleton. *Int J Biochem Cell Biol,* 40**,** 46-62.

Maeda, N. 2015. Proteoglycans and neuronal migration in the cerebral cortex during development and disease. *Front Neurosci,* 9**,** 98.

Maehara, H., Suzuki, K., Sasaki, T., Oshita, H., Wada, E., Inoue, T. & Shimizu, K. 2007. G1-G2 Aggrecan Product that can be Generated by M-calpain on Truncation at Ala709–Ala710 is Present Abundantly in Human Articular Cartilage. *Journal of Biochemistry,* 141**,** 469-477.

Mak, K. K., Kronenberg, H. M., Chuang, P. T., Mackem, S. & Yang, Y. 2008. Indian hedgehog signals independently of PTHrP to promote chondrocyte hypertrophy. *Development,* 135**,** 1947-56.

Maroudas, A., Bayliss, M. T., Uchitel-Kaushansky, N., Schneiderman, R. & Gilav, E. 1998. Aggrecan turnover in human articular cartilage: use of aspartic acid racemization as a marker of molecular age. *Arch Biochem Biophys,* 350**,** 61-71.

Maston, G. A., Evans, S. K. & Green, M. R. 2006. Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet,* 7**,** 29-59.

Matsumoto, K., Shionyu, M., Go, M., Shimizu, K., Shinomura, T., Kimata, K. & Watanabe, H. 2003. Distinct interaction of versican/PG-M with hyaluronan and link protein. *J Biol Chem,* 278**,** 41205-12.

Matys, V., Kel-Margoulis, O. V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K., Voss, N., Stegmaier, P., Lewicki-Potapov, B., Saxel, H., Kel, A. E. & Wingender, E. 2006. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res,* 34**,** D108-10.

Mcalinden, A., Havlioglu, N., Liang, L., Davies, S. R. & Sandell, L. J. 2005. Alternative splicing of type II procollagen exon 2 is regulated by the combination of a weak 5' splice site and an adjacent intronic stem-loop cis element. *J Biol Chem,* 280**,** 32700-11.

Mccann, M. R., Tamplin, O. J., Rossant, J. & Séguin, C. A. 2012. Tracing notochord-derived cells using a Noto-cre mouse: implications for intervertebral disc development. *Disease Models and Mechanisms,* 5**,** 73-82.

Mckeown-Longo, P. J. & Goetinck, P. F. 1982. Characterization of the tissue-specific proteoglycans synthesized by chondrocytes from nanomelic chick embryos. *Biochem J,* 201**,** 387-94.

Mead, T. J., Wang, Q., Bhattaram, P., Dy, P., Afelik, S., Jensen, J. & Lefebvre, V. 2013. A far-upstream (-70 kb) enhancer mediates Sox9 auto-regulation in somatic tissues during development and adult regeneration. *Nucleic Acids Res,* 41**,** 4459-69.

Mead, T. J. & Yutzey, K. E. 2009. Notch pathway regulation of chondrocyte differentiation and proliferation during appendicular and axial skeleton development. *Proc Natl Acad Sci U S A,* 106**,** 14420-5.

Melin Furst, C., Morgelin, M., Vadstrup, K., Heinegard, D., Aspberg, A. & Blom, A. M. 2013. The C-type lectin of the aggrecan G3 domain activates complement. *PLoS One,* 8**,** e61407.

Melrose, J., Smith, S. M., Appleyard, R. C. & Little, C. B. 2008. Aggrecan, versican and type VI collagen are components of annular translamellar crossbridges in the intervertebral disc. *Eur Spine J,* 17**,** 314-24.

Menzel, U., Kosteas, T., Tolaini, M., Killeen, N., Roderick, K. & Kioussis, D. 2011. Modulation of the murine CD8 gene complex following the targeted integration of human CD2-locus control region sequences. *J Immunol,* 187**,** 3712-20.

Metzger, B. P., Yuan, D. C., Gruber, J. D., Duveau, F. & Wittkopp, P. J. 2015. Selection on noise constrains variation in a eukaryotic promoter. *Nature,* 521**,** 344-7.

Meyer, C. A. & Liu, X. S. 2014. Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. *Nat Rev Genet,* 15**,** 709-21.

Meyer, M. B., Benkusky, N. A. & Pike, J. W. 2015. Selective Distal Enhancer Control of the Mmp13 Gene Identified through Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) Genomic Deletions. *J Biol Chem,* 290**,** 11093-107.

Milz, S. & Putz, R. 1994. Quantitative morphology of the subchondral plate of the tibial plateau. *Journal of Anatomy,* 185**,** 103-110.

Minguillon, C., Nishimoto, S., Wood, S., Vendrell, E., Gibson-Brown, J. J. & Logan, M. P. 2012. Hox genes regulate the onset of Tbx5 expression in the forelimb. *Development,* 139**,** 3180-8.

Mirando, A. J., Liu, Z., Moore, T., Lang, A., Kohn, A., Osinski, A. M., O'keefe, R. J., Mooney, R. A., Zuscik, M. J. & Hilton, M. J. 2013. RBP-Jkappa-dependent Notch signaling is required for murine articular cartilage and joint maintenance. *Arthritis Rheum,* 65**,** 2623-33.

Miura, H., Gurumurthy, C. B., Sato, T., Sato, M. & Ohtsuka, M. 2015. CRISPR/Cas9-based generation of knockdown mice by intronic insertion of artificial microRNA using longer single-stranded DNA. *Scientific Reports,* 5**,** 12799.

Morawski, M., Bruckner, G., Arendt, T. & Matthews, R. T. 2012. Aggrecan: Beyond cartilage and into the brain. *Int J Biochem Cell Biol,* 44**,** 690-3.

Muhr, J., Andersson, E., Persson, M., Jessell, T. M. & Ericson, J. 2001. Groucho-Mediated Transcriptional Repression Establishes Progenitor Cell Pattern and Neuronal Fate in the Ventral Neural Tube. *Cell,* 104**,** 861-873.

Murakami, K., Calero, G., Brown, C. R., Liu, X., Davis, R. E., Boeger, H. & Kornberg, R. D. 2013. Formation and fate of a complete 31-protein RNA polymerase II transcription preinitiation complex. *J Biol Chem,* 288**,** 6325-32.

Nakajima, M., Takahashi, A., Kou, I., Rodriguez-Fontenla, C., Gomez-Reino, J. J., Furuichi, T., Dai, J., Sudo, A., Uchida, A., Fukui, N., Kubo, M., Kamatani, N., Tsunoda, T., Malizos, K. N., Tsezou, A., Gonzalez, A., Nakamura, Y. & Ikegawa, S. 2010. New sequence variants in HLA class II/III region associated with susceptibility to knee osteoarthritis identified by genome-wide association study. *PLoS One,* 5.

Nakamura, E., Nguyen, M. T. & Mackem, S. 2006. Kinetics of tamoxifen-regulated Cre activity in mice using a cartilage-specific CreER(T) to assay temporal activity windows along the proximodistal limb skeleton. *Dev Dyn,* 235**,** 2603-12.

Naso, M. F., Morgan, J. L., Buchberg, A. M., Siracusa, L. D. & Iozzo, R. V. 1995. Expression pattern and mapping of the murine versican gene (Cspg2) to chromosome 13. *Genomics,* 29**,** 297-300.

Neidlinger-Wilke, C., Wurtz, K., Urban, J. P., Borm, W., Arand, M., Ignatius, A., Wilke, H. J. & Claes, L. E. 2006. Regulation of gene expression in intervertebral disc cells by low and high hydrostatic pressure. *Eur Spine J,* 15 Suppl 3**,** S372-8.

Nickol, J. M. & Felsenfeld, G. 1988. Bidirectional control of the chicken beta- and epsilon-globin genes by a shared enhancer. *Proc Natl Acad Sci U S A,* 85**,** 2548-52.

Nilsson, O., Guo, M. H., Dunbar, N., Popovic, J., Flynn, D., Jacobsen, C., Lui, J. C., Hirschhorn, J. N., Baron, J. & Dauber, A. 2014. Short stature, accelerated bone maturation, and early growth cessation due to heterozygous aggrecan mutations. *J Clin Endocrinol Metab,* 99**,** E1510-8.

Nissim, L., Perli, S. D., Fridkin, A., Perez-Pinera, P. & Lu, T. K. 2014. Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells. *Mol Cell,* 54**,** 698-710.

Nowlan, N. C., Sharpe, J., Roddy, K. A., Prendergast, P. J. & Murphy, P. 2010. Mechanobiology of embryonic skeletal development: Insights from animal models. *Birth Defects Res C Embryo Today,* 90**,** 203-13.

Ochi, H., Tamai, T., Nagano, H., Kawaguchi, A., Sudou, N. & Ogino, H. 2012. Evolution of a tissue-specific silencer underlies divergence in the expression of pax2 and pax8 paralogues. *Nat Commun,* 3**,** 848.

Oh, C. D., Yasuda, H., Zhao, W., Henry, S. P., Zhang, Z., Xue, M., De Crombrugghe, B. & Chen, D. 2016. SOX9 directly Regulates CTGF/CCN2 Transcription in Growth Plate Chondrocytes and in Nucleus Pulposus Cells of Intervertebral Disc. *Sci Rep,* 6**,** 29916.

Ohba, S., He, X., Hojo, H. & Mcmahon, Andrew p. 2015. Distinct Transcriptional Programs Underlie Sox9 Regulation of the Mammalian Chondrocyte. *Cell Reports*.

Olins, A. L. & Olins, D. E. 1974. Spheroid chromatin units (v bodies). *Science,* 183**,** 330-2.

Ong, C. T. & Corces, V. G. 2011. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat Rev Genet,* 12**,** 283-93.

Ong, S. J., Huang, L. C., Liu, H. W., Chang, S. C., Yang, Y. C., Bessarab, I. & Tai, J. H. 2002. Characterization of a bi-directional promoter for divergent transcription of a PHD-zinc finger protein gene and a ran gene in the protozoan pathogen Giardia lamblia. *Mol Microbiol,* 43**,** 665-76.

Ono, N., Ono, W., Nagasawa, T. & Kronenberg, H. M. 2014. A subset of chondrogenic cells provides early mesenchymal progenitors in growing bones. *Nat Cell Biol,* 16**,** 1157-67.

Ovchinnikov, D. A., Deng, J. M., Ogunrinu, G. & Behringer, R. R. 2000. Col2a1-directed expression of Cre recombinase in differentiating chondrocytes in transgenic mice. *genesis,* 26**,** 145-146.

Pacifici, M., Koyama, E., Shibukawa, Y., Wu, C., Tamamura, Y., Enomoto-Iwamoto, M. & Iwamoto, M. 2006. Cellular and molecular mechanisms of synovial joint and articular cartilage formation. *Ann N Y Acad Sci,* 1068**,** 74-86.

Pandey, R. & Dou, Y. 2013. H2A.Z sets the stage in ESCs. *Cell Stem Cell,* 12**,** 143-4.

Pantazopoulos, H., Markota, M., Jaquet, F., Ghosh, D., Wallin, A., Santos, A., Caterson, B. & Berretta, S. 2015. Aggrecan and chondroitin-6-sulfate abnormalities in schizophrenia and bipolar disorder: a postmortem study on the amygdala. *Transl Psychiatry,* 5**,** e496.

Parker, S. C., Stitzel, M. L., Taylor, D. L., Orozco, J. M., Erdos, M. R., Akiyama, J. A., Van Bueren, K. L., Chines, P. S., Narisu, N., Program, N. C. S., Black, B. L., Visel, A., Pennacchio, L. A., Collins, F. S., National Institutes of Health Intramural Sequencing Center Comparative Sequencing Program, A. & Authors, N. C. S. P. 2013. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc Natl Acad Sci U S A,* 110**,** 17921-6.

Pefanis, E., Wang, J., Rothschild, G., Lim, J., Kazadi, D., Sun, J., Federation, A., Chao, J., Elliott, O., Liu, Z. P., Economides, A. N., Bradner, J. E., Rabadan, R. & Basu, U. 2015. RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. *Cell,* 161**,** 774-89.

Pekowska, A., Benoukraf, T., Zacarias-Cabeza, J., Belhocine, M., Koch, F., Holota, H., Imbert, J., Andrau, J. C., Ferrier, P. & Spicuglia, S. 2011. H3K4 tri-methylation provides an epigenetic signature of active enhancers. *Embo j,* 30**,** 4198-210.

Pennacchio, L. A., Ahituv, N., Moses, A. M., Prabhakar, S., Nobrega, M. A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K. D., Plajzer-Frick, I., Akiyama, J., De

Val, S., Afzal, V., Black, B. L., Couronne, O., Eisen, M. B., Visel, A. & Rubin, E. M. 2006. In vivo enhancer analysis of human conserved non-coding sequences. *Nature,* 444**,** 499-502.

Percival, C. J. & Richtsmeier, J. T. 2013. Angiogenesis and intramembranous osteogenesis. *Dev Dyn,* 242**,** 909-22.

Perry, M. W., Boettiger, A. N., Bothma, J. P. & Levine, M. 2010. Shadow enhancers foster robustness of Drosophila gastrulation. *Curr Biol,* 20**,** 1562-7.

Pirok, E. W., Li, H., Mensch, J. R., Henry, J. & Schwartz, N. B. 1997. Structural and Functional Analysis of the Chick Chondroitin Sulfate Proteoglycan (Aggrecan) Promoter and Enhancer Region. *Journal of Biological Chemistry,* 272**,** 11566-11574.

Pitsillides, A. A. & Ashhurst, D. E. 2008. A critical evaluation of specific aspects of joint development. *Dev Dyn,* 237**,** 2284-94.

Plaschka, C., Lariviere, L., Wenzeck, L., Seizl, M., Hemann, M., Tegunov, D., Petrotchenko, E. V., Borchers, C. H., Baumeister, W., Herzog, F., Villa, E. & Cramer, P. 2015. Architecture of the RNA polymerase II-Mediator core initiation complex. *Nature,* 518**,** 376-80.

Pott, S. & Lieb, J. D. 2014. What are super-enhancers? *Nat Genet,* 47**,** 8-12.

Poulet, B. 2016. Non-invasive Loading Model of Murine Osteoarthritis. *Curr Rheumatol Rep,* 18**,** 40.

Poulet, B., Hamilton, R. W., Shefelbine, S. & Pitsillides, A. A. 2011. Characterizing a novel and adjustable noninvasive murine joint loading model. *Arthritis Rheum,* 63**,** 137-47.

Pradeepa, M. M., Grimes, G. R., Kumar, Y., Olley, G., Taylor, G. C., Schneider, R. & Bickmore, W. A. 2016. Histone H3 globular domain acetylation identifies a new class of enhancers. *Nat Genet*.

Proudhon, C., Snetkova, V., Raviram, R., Lobry, C., Badri, S., Jiang, T., Hao, B., Trimarchi, T., Kluger, Y., Aifantis, I., Bonneau, R. & Skok, Jane a. 2016. Active and Inactive

Enhancers Cooperate to Exert Localized and Long-Range Control of Gene Regulation. *Cell Reports*.

Quang, D. X., Erdos, M. R., Parker, S. C. & Collins, F. S. 2015. Motif signatures in stretch enhancers are enriched for disease-associated genetic variants. *Epigenetics Chromatin,* 8**,** 23.

Radons, J., Bosserhoff, A. K., Grassel, S., Falk, W. & Schubert, T. E. 2006. p38MAPK mediates IL-1-induced down-regulation of aggrecan gene expression in human chondrocytes. *Int J Mol Med,* 17**,** 661-8.

Ran, F. A., Cong, L., Yan, W. X., Scott, D. A., Gootenberg, J. S., Kriz, A. J., Zetsche, B., Shalem, O., Wu, X., Makarova, K. S., Koonin, E. V., Sharp, P. A. & Zhang, F. 2015. In vivo genome editing using Staphylococcus aureus Cas9. *Nature,* 520**,** 186-91.

Ran, F. A., Hsu, P. D., Lin, C. Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., Scott, D. A., Inoue, A., Matoba, S., Zhang, Y. & Zhang, F. 2013. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell,* 154**,** 1380-9.

Rasmussen, S., Glickman, G., Norinsky, R., Quimby, F. W. & Tolwani, R. J. 2009. Construction Noise Decreases Reproductive Efficiency in Mice. *Journal of the American Association for Laboratory Animal Science : JAALAS,* 48**,** 363-370.

Reesink, H. L., Bonnevie, E. D., Liu, S., Shurer, C. R., Hollander, M. J., Bonassar, L. J. & Nixon, A. J. 2016. Galectin-3 Binds to Lubricin and Reinforces the Lubricating Boundary Layer of Articular Cartilage. *Sci Rep,* 6**,** 25463.

Retting, K. N., Song, B., Yoon, B. S. & Lyons, K. M. 2009. BMP canonical Smad signaling through Smad1 and Smad5 is required for endochondral bone formation. *Development,* 136**,** 1093-104.

Reynard, L. N. 2016. Analysis of genetics and DNA methylation in osteoarthritis: What have we learnt about the disease? *Seminars in Cell & Developmental Biology*.

Reynard, L. N., Bui, C., Syddall, C. M. & Loughlin, J. 2014. CpG methylation regulates allelic expression of GDF5 by modulating binding of SP1 and SP3 repressor proteins to the osteoarthritis susceptibility SNP rs143383. *Hum Genet,* 133**,** 1059-73.

Reynolds, R. P., Kinard, W. L., Degraff, J. J., Leverage, N. & Norton, J. N. 2010. Noise in a Laboratory Animal Facility from the Human and Mouse Perspectives. *Journal of the American Association for Laboratory Animal Science : JAALAS,* 49**,** 592-597.

Rice, D. P., Aberg, T., Chan, Y., Tang, Z., Kettunen, P. J., Pakarinen, L., Maxson, R. E. & Thesleff, I. 2000. Integration of FGF and TWIST in calvarial bone and suture development. *Development,* 127**,** 1845-55.

Riso, V., Cammisa, M., Kukreja, H., Anvar, Z., Verde, G., Sparago, A., Acurzio, B., Lad, S., Lonardo, E., Sankar, A., Helin, K., Feil, R., Fico, A., Angelini, C., Grimaldi, G. & Riccio, A. 2016. ZFP57 maintains the parent-of-origin-specific expression of the imprinted genes and differentially affects non-imprinted targets in mouse embryonic stem cells. *Nucleic Acids Research*.

Ristolainen, H., Kilpivaara, O., Kamper, P., Taskinen, M., Saarinen, S., Leppä, S., D'amore, F. & Aaltonen, L. A. 2015. Identification of homozygous deletion in ACAN and other candidate variants in familial classical Hodgkin lymphoma by exome sequencing. *British Journal of Haematology,* 170**,** 428-431.

Robinson, P. J., Trnka, M. J., Bushnell, D. A., Davis, R. E., Mattei, P. J., Burlingame, A. L. & Kornberg, R. D. 2016. Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. *Cell,* 166**,** 1411-1422 e16.

Rodova, M., Lu, Q., Li, Y., Woodbury, B. G., Crist, J. D., Gardner, B. M., Yost, J. G., Zhong, X. B., Anderson, H. C. & Wang, J. 2011. Nfat1 regulates adult articular chondrocyte function through its age-dependent expression mediated by epigenetic histone methylation. *J Bone Miner Res,* 26**,** 1974-86.

Rodrigo, I., Hill, R. E., Balling, R., Munsterberg, A. & Imai, K. 2003. Pax1 and Pax9 activate Bapx1 to induce chondrogenic differentiation in the sclerotome. *Development,* 130**,** 473-82.

Roeder, R. G. & Rutter, W. J. 1969. Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature,* 224**,** 234-7.

Rogerson, F. M., Chung, Y. M., Deutscher, M. E., Last, K. & Fosang, A. J. 2010. Cytokine-induced increases in ADAMTS-4 messenger RNA expression do not lead to

increased aggrecanase activity in ADAMTS-5-deficient mice. *Arthritis Rheum,* 62**,** 3365-73.

Rojas, A., De Val, S., Heidt, A. B., Xu, S. M., Bristow, J. & Black, B. L. 2005. Gata4 expression in lateral mesoderm is downstream of BMP4 and is activated directly by Forkhead and GATA transcription factors through a distal enhancer element. *Development,* 132**,** 3405-17.

Rossi, P. & De Crombrugghe, B. 1987. Identification of a cell-specific transcriptional enhancer in the first intron of the mouse alpha 2 (type I) collagen gene. *Proc Natl Acad Sci U S A,* 84**,** 5590-4.

Rothbart, S. B. & Strahl, B. D. 2014. Interpreting the language of histone and DNA modifications. *Biochim Biophys Acta,* 1839**,** 627-43.

Roughley, P. J. 2006. The structure and function of cartilage proteoglycans. *Eur Cell Mater,* 12**,** 92-101.

Rountree, R. B., Schoor, M., Chen, H., Marks, M. E., Harley, V., Mishina, Y. & Kingsley, D. M. 2004. BMP receptor signaling is required for postnatal maintenance of articular cartilage. *PLoS Biol,* 2**,** e355.

Rudnicki, M., Perco, P., Neuwirt, H., Noppert, S.-J., Leierer, J., Sunzenauer, J., Eder, S., Zoja, C., Eller, K., Rosenkranz, A. R., Müller, G. A., Mayer, B. & Mayer, G. 2012. Increased Renal Versican Expression Is Associated with Progression of Chronic Kidney Disease. *PLoS ONE,* 7**,** e44891.

Rushton, M. D., Reynard, L. N., Barter, M. J., Refaie, R., Rankin, K. S., Young, D. A. & Loughlin, J. 2014. Characterization of the Cartilage DNA Methylome in Knee and Hip Osteoarthritis. *Arthritis & Rheumatology,* 66**,** 2450-2460.

Sacilotto, N., Monteiro, R., Fritzsche, M., Becker, P. W., Sanchez-Del-Campo, L., Liu, K., Pinheiro, P., Ratnayaka, I., Davies, B., Goding, C. R., Patient, R., Bou-Gharios, G. & De Val, S. 2013. Analysis of Dll4 regulation reveals a combinatorial role for Sox and Notch in arterial development. *Proc Natl Acad Sci U S A,* 110**,** 11893-8.

Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M. & Shiroishi, T. 2005. Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development,* 132**,** 797-803.

Sander, J. D. & Joung, J. K. 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol,* 32**,** 347-55.

Sandy, J. D., Flannery, C. R., Neame, P. J. & Lohmander, L. S. 1992. The structure of aggrecan fragments in human synovial fluid. Evidence for the involvement in osteoarthritis of a novel proteinase which cleaves the Glu 373-Ala 374 bond of the interglobular domain. *The Journal of Clinical Investigation,* 89**,** 1512-1516.

Sarcinella, E., Zuzarte, P. C., Lau, P. N., Draker, R. & Cheung, P. 2007. Monoubiquitylation of H2A.Z distinguishes its association with euchromatin or facultative heterochromatin. *Mol Cell Biol,* 27**,** 6457-68.

Sato, T., Kudo, T., Ikehara, Y., Ogawa, H., Hirano, T., Kiyohara, K., Hagiwara, K., Togayachi, A., Ema, M., Takahashi, S., Kimata, K., Watanabe, H. & Narimatsu, H. 2011. Chondroitin sulfate N-acetylgalactosaminyltransferase 1 is necessary for normal endochondral ossification and aggrecan metabolism. *J Biol Chem,* 286**,** 5803-12.

Sauer, B. & Henderson, N. 1988. Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. *Proceedings of the National Academy of Sciences of the United States of America,* 85**,** 5166-5170.

Savagner, P., Miyashita, T. & Yamada, Y. 1990. Two silencers regulate the tissue-specific expression of the collagen II gene. *Journal of Biological Chemistry,* 265**,** 6669-74.

Schubeler, D., Groudine, M. & Bender, M. A. 2001. The murine beta-globin locus control region regulates the rate of transcription but not the hyperacetylation of histones at the active genes. *Proc Natl Acad Sci U S A,* 98**,** 11432-7.

Schwartz, N. B. & Domowicz, M. 2002. Chondrodysplasias due to proteoglycan defects. *Glycobiology,* 12**,** 57R-68R.

Schwertman, P., Bekker-Jensen, S. & Mailand, N. 2016. Regulation of DNA double-strand break repair by ubiquitin and ubiquitin-like modifiers. *Nat Rev Mol Cell Biol,* 17**,** 379-394.

Sebinger, D. D., Ofenbauer, A., Gruber, P., Malik, S. & Werner, C. 2013. ECM modulated early kidney development in embryonic organ culture. *Biomaterials,* 34**,** 6670-82.

Seki, K., Fujimori, T., Savagner, P., Hata, A., Aikawa, T., Ogata, N., Nabeshima, Y. & Kaechoong, L. 2003. Mouse Snail family transcription repressors regulate chondrocyte, extracellular matrix, type II collagen, and aggrecan. *J Biol Chem,* 278**,** 41862-70.

Sekiya, I., Tsuji, K., Koopman, P., Watanabe, H., Yamada, Y., Shinomiya, K., Nifuji, A. & Noda, M. 2000. SOX9 enhances aggrecan gene promoter/enhancer activity and is up-regulated by retinoic acid in a cartilage-derived cell line, TC6. *J Biol Chem,* 275**,** 10738-44.

Selvaraj, S., R Dixon, J., Bansal, V. & Ren, B. 2013. Whole-genome haplotype reconstruction using proximity-ligation and shotgun sequencing. *Nat Biotech,* 31**,** 1111-1118.

Shalem, O., Sanjana, N. E. & Zhang, F. 2015. High-throughput functional genomics using CRISPR-Cas9. *Nat Rev Genet,* 16**,** 299-311.

Sharma, S. & Zhu, J. 2014. Immunologic Applications of Conditional Gene Modification Technology in the Mouse. *Current protocols in immunology / edited by John E. Coligan ... [et al.],* 105**,** 10.34.1-10.34.13.

Shema, E., Jones, D., Shoresh, N., Donohue, L., Ram, O. & Bernstein, B. E. 2016. Single-molecule decoding of combinatorially modified nucleosomes. *Science,* 352**,** 717-721.

Shen, Y., Yue, F., Mccleary, D. F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V. V. & Ren, B. 2012. A map of the cis-regulatory sequences in the mouse genome. *Nature,* 488**,** 116-20.

Shi, S., Grothe, S., Zhang, Y., O'connor-Mccourt, M. D., Poole, A. R., Roughley, P. J. & Mort, J. S. 2004. Link protein has greater affinity for versican than aggrecan. *J Biol Chem,* 279**,** 12060-6.

Shibata, S., Fukada, K., Imai, H., Abe, T. & Yamashita, Y. 2003. In situ hybridization and immunohistochemistry of versican, aggrecan and link protein, and histochemistry of hyaluronan in the developing mouse limb bud cartilage. *J Anat,* 203**,** 425-32.

Shinomura, T., Ito, K., Hook, M. & Kimura, J. H. 2012. A newly identified enhancer element responsible for type II collagen gene expression. *J Biochem,* 152**,** 565-75.

Shinomura, T., Ito, K., Kimura, J. H. & Höök, M. 2006. Screening for genes preferentially expressed in the early phase of chondrogenesis. *Biochemical and Biophysical Research Communications,* 341**,** 167-174.

Shukunami, C., Ishizeki, K., Atsumi, T., Ohta, Y., Suzuki, F. & Hiraki, Y. 1997. Cellular hypertrophy and calcification of embryonal carcinoma-derived chondrogenic cell line ATDC5 in vitro. *J Bone Miner Res,* 12**,** 1174-88.

Siegal, M. L. & Bergman, A. 2002. Waddington's canalization revisited: developmental stability and evolution. *Proc Natl Acad Sci U S A,* 99**,** 10528-32.

Siggens, L. & Ekwall, K. 2014. Epigenetics, chromatin and genome organization: recent advances from the ENCODE project. *J Intern Med,* 276**,** 201-14.

Smale, S. T. & Baltimore, D. 1989. The "initiator" as a transcription control element. *Cell,* 57**,** 103-13.

Soboleva, T. A., Nekrasov, M., Ryan, D. P. & Tremethick, D. J. 2014. Histone variants at the transcription start-site. *Trends Genet*.

Sohaskey, M. L., Yu, J., Diaz, M. A., Plaas, A. H. & Harland, R. M. 2008. JAWS coordinates chondrogenesis and synovial joint positioning. *Development,* 135**,** 2215-2220.

Sophia Fox, A. J., Bedi, A. & Rodeo, S. A. 2009. The Basic Science of Articular Cartilage: Structure, Composition, and Function. *Sports Health,* 1**,** 461-468.

Soriano, P. 1999. Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nat Genet,* 21**,** 70-71.

Soucie, E. L., Weng, Z., Geirsdóttir, L., Molawi, K., Maurizio, J., Fenouil, R., Mossadegh-Keller, N., Gimenez, G., Vanhille, L., Beniazza, M., Favret, J., Berruyer, C., Perrin, P., Hacohen, N., Andrau, J.-C., Ferrier, P., Dubreuil, P., Sidow, A. & Sieweke, M. H. 2016. Lineage-specific enhancers activate self-renewal genes in macrophages and embryonic stem cells. *Science*.

Soutoglou, E., Katrakili, N. & Talianidis, I. 2000. Acetylation Regulates Transcription Factor Activity at Multiple Levels. *Molecular Cell,* 5**,** 745-751.

Spicer, A. P., Joo, A. & Bowling, R. A., Jr. 2003. A hyaluronan binding link protein gene family whose members are physically linked adjacent to chondroitin sulfate proteoglycan core protein genes: the missing links. *J Biol Chem,* 278**,** 21083-91.

Spitz, F., Gonzalez, F. & Duboule, D. 2003. A Global Control Region Defines a Chromosomal Regulatory Landscape Containing the HoxD Cluster. *Cell,* 113**,** 405-417.

Spruijt, C. G. & Vermeulen, M. 2014. DNA methylation: old dog, new tricks[quest]. *Nat Struct Mol Biol,* 21**,** 949-954.

Staller, M. V., Vincent, B. J., Bragdon, M. D. J., Lydiard-Martin, T., Wunderlich, Z., Estrada, J. & Depace, A. H. 2015. Shadow enhancers enable Hunchback bifunctionality in the Drosophila embryo. *Proceedings of the National Academy of Sciences,* 112**,** 785-790.

Stanton, H., Melrose, J., Little, C. B. & Fosang, A. J. 2011. Proteoglycan degradation by the ADAMTS family of proteinases. *Biochim Biophys Acta,* 1812**,** 1616-29.

Stattin, E. L., Wiklund, F., Lindblom, K., Onnerfjord, P., Jonsson, B. A., Tegner, Y., Sasaki, T., Struglics, A., Lohmander, S., Dahl, N., Heinegard, D. & Aspberg, A. 2010. A missense mutation in the aggrecan C-type lectin domain disrupts extracellular matrix interactions and causes dominant familial osteochondritis dissecans. *Am J Hum Genet,* 86**,** 126-37.

Stein, P. & Schindler, K. 2011. Mouse Oocyte Microinjection, Maturation and Ploidy Assessment. e2851.

Struglics, A. & Hansson, M. 2010. Calpain is involved in C-terminal truncation of human aggrecan. *Biochem J,* 430**,** 531-8.

Struglics, A. & Hansson, M. 2012. MMP proteolysis of the human extracellular matrix protein aggrecan is mainly a process of normal turnover. *Biochem J,* 446**,** 213-23.

Subramanian, V., Mazumder, A., Surface, L. E., Butty, V. L., Fields, P. A., Alwan, A., Torrey, L., Thai, K. K., Levine, S. S., Bathe, M. & Boyer, L. A. 2013. H2A.Z acidic patch couples chromatin dynamics to regulation of gene expression programs during ESC differentiation. *PLoS Genet,* 9**,** e1003725.

Sugimoto, Y., Takimoto, A., Akiyama, H., Kist, R., Scherer, G., Nakamura, T., Hiraki, Y. & Shukunami, C. 2013a. Scx+/Sox9+ progenitors contribute to the establishment of the junction between cartilage and tendon/ligament. *Development,* 140**,** 2280-8.

Sugimoto, Y., Takimoto, A., Hiraki, Y. & Shukunami, C. 2013b. Generation and characterization of ScxCre transgenic mice. *genesis,* 51**,** 275-283.

Sun, F.-L. & Elgin, S. C. R. 1999. Putting Boundaries on Silence. *Cell,* 99**,** 459-462.

Suwanwela, J., Farber, C. R., Haung, B.-L., Song, B., Pan, C., Lyons, K. M. & Lusis, A. J. 2011. Systems Genetics Analysis of Mouse Chondrocyte Differentiation. *Journal of Bone and Mineral Research,* 26**,** 747-760.

Tagami, H., Ray-Gallet, D., Almouzni, G. & Nakatani, Y. 2004. Histone H3.1 and H3.3 Complexes Mediate Nucleosome Assembly Pathways Dependent or Independent of DNA Synthesis. *Cell,* 116**,** 51-61.

Taher, L., Collette, N. M., Murugesh, D., Maxwell, E., Ovcharenko, I. & Loots, G. G. 2011. Global Gene Expression Analysis of Murine Limb Development. *PLoS ONE,* 6**,** e28358.

Takagi, Y. & Kornberg, R. D. 2006. Mediator as a general transcription factor. *J Biol Chem,* 281**,** 80-9.

Takimoto, A., Mohri, H., Kokubu, C., Hiraki, Y. & Shukunami, C. 2013. Pax1 acts as a negative regulator of chondrocyte maturation. *Exp Cell Res,* 319**,** 3128-39.

Talbot, D., Collis, P., Antoniou, M., Vidal, M., Grosveld, F. & Greaves, D. R. 1989. A dominant control region from the human beta-globin locus conferring integration site-independent gene expression. *Nature,* 338**,** 352-5.

Tan, M., Luo, H., Lee, S., Jin, F., Yang, J. S., Montellier, E., Buchou, T., Cheng, Z., Rousseaux, S., Rajagopal, N., Lu, Z., Ye, Z., Zhu, Q., Wysocka, J., Ye, Y., Khochbin, S., Ren, B. & Zhao, Y. 2011. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell,* 146**,** 1016-28.

Tan, Y., Xue, Y., Song, C. & Grunstein, M. 2013. Acetylated histone H3K56 interacts with Oct4 to promote mouse embryonic stem cell pluripotency. *Proc Natl Acad Sci U S A,* 110**,** 11493-8.

Teng, L., He, B., Wang, J. & Tan, K. 2015. 4DGenome: a comprehensive database of chromatin interactions. *Bioinformatics,* 31**,** 2560-4.

Tenu, J.-P., Viratelle, O. M., Garnier, J. & Yon, J. 1971. pH Dependence of the Activity of β-Galactosidase from Escherichia coli. *European Journal of Biochemistry,* 20**,** 363-370.

Tessarz, P. & Kouzarides, T. 2014. Histone core modifications regulating nucleosome structure and dynamics. *Nat Rev Mol Cell Biol,* 15**,** 703-8.

Thakore, P. I., D'ippolito, A. M., Song, L., Safi, A., Shivakumar, N. K., Kabadi, A. M., Reddy, T. E., Crawford, G. E. & Gersbach, C. A. 2015. Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat Methods,* 12**,** 1143-9.

Thomas-Chollier, M., Hufton, A., Heinig, M., O'keeffe, S., Masri, N. E., Roider, H. G., Manke, T. & Vingron, M. 2011. Transcription factor binding predictions using TRAP for the analysis of ChIP-seq data and regulatory SNPs. *Nat Protoc,* 6**,** 1860-9.

Thomas, M. C. & Chiang, C. M. 2006. The general transcription machinery and general cofactors. *Crit Rev Biochem Mol Biol,* 41**,** 105-78.

Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T., Haugen, E., Sheffield, N. C., Stergachis, A. B., Wang, H., Vernot, B., Garg, K., John, S.,

Sandstrom, R., Bates, D., Boatman, L., Canfield, T. K., Diegel, M., Dunn, D., Ebersol, A. K., Frum, T., Giste, E., Johnson, A. K., Johnson, E. M., Kutyavin, T., Lajoie, B., Lee, B. K., Lee, K., London, D., Lotakis, D., Neph, S., Neri, F., Nguyen, E. D., Qu, H., Reynolds, A. P., Roach, V., Safi, A., Sanchez, M. E., Sanyal, A., Shafer, A., Simon, J. M., Song, L., Vong, S., Weaver, M., Yan, Y., Zhang, Z., Zhang, Z., Lenhard, B., Tewari, M., Dorschner, M. O., Hansen, R. S., Navas, P. A., Stamatoyannopoulos, G., Iyer, V. R., Lieb, J. D., Sunyaev, S. R., Akey, J. M., Sabo, P. J., Kaul, R., Furey, T. S., Dekker, J., Crawford, G. E. & Stamatoyannopoulos, J. A. 2012. The accessible chromatin landscape of the human genome. *Nature,* 489**,** 75-82.

Tiecke, E., Bangs, F., Blaschke, R., Farrell, E. R., Rappold, G. & Tickle, C. 2006. Expression of the short stature homeobox gene Shox is restricted by proximal and distal signals in chick limb buds and affects the length of skeletal elements. *Dev Biol,* 298**,** 585-96.

Tolhuis, B., Palstra, R.-J., Splinter, E., Grosveld, F. & De Laat, W. 2002. Looping and Interaction between Hypersensitive Sites in the Active β-globin Locus. *Molecular Cell,* 10**,** 1453-1465.

Tompson, S. W., Merriman, B., Funari, V. A., Fresquet, M., Lachman, R. S., Rimoin, D. L., Nelson, S. F., Briggs, M. D., Cohn, D. H. & Krakow, D. 2009. A recessive skeletal dysplasia, SEMD aggrecan type, results from a missense mutation affecting the C-type lectin domain of aggrecan. *Am J Hum Genet,* 84**,** 72-9.

Tran, T. H., Jarrell, A., Zentner, G. E., Welsh, A., Brownell, I., Scacheri, P. C. & Atit, R. 2010. Role of canonical Wnt signaling/ss-catenin via Dermo1 in cranial dermal cell development. *Development,* 137**,** 3973-84.

Trinklein, N. D., Aldred, S. F., Hartman, S. J., Schroeder, D. I., Otillar, R. P. & Myers, R. M. 2004. An Abundance of Bidirectional Promoters in the Human Genome. *Genome Research,* 14**,** 62-66.

Troeberg, L., Lazenbatt, C., Anower, E. K. M. F., Freeman, C., Federov, O., Habuchi, H., Habuchi, O., Kimata, K. & Nagase, H. 2014. Sulfated Glycosaminoglycans Control

the Extracellular Trafficking and the Activity of the Metalloprotease Inhibitor TIMP-3. *Chem Biol*.

Troeberg, L. & Nagase, H. 2012. Proteases involved in cartilage matrix degradation in osteoarthritis. *Biochim Biophys Acta,* 1824**,** 133-45.

Trojer, P. & Reinberg, D. 2007. Facultative heterochromatin: is there a distinctive molecular signature? *Mol Cell,* 28**,** 1-13.

Tsai, Y.-C., Cooke, N. E. & Liebhaber, S. A. 2016. Long-range looping of a locus control region drives tissue-specific chromatin packing within a multigene cluster. *Nucleic Acids Research,* 44**,** 4651-4664.

Tsang, K. Y., Chan, D. & Cheah, K. S. 2015. Fate of growth plate hypertrophic chondrocytes: death or lineage extension? *Dev Growth Differ,* 57**,** 179-92.

Tschopp, P., Fraudeau, N., Béna, F. & Duboule, D. 2011. Reshuffling genomic landscapes to study the regulatory evolution of Hox gene clusters. *Proceedings of the National Academy of Sciences,* 108**,** 10632-10637.

Tzanoulinou, S., Brandi, R., Arisi, I., D'onofrio, M., Urfer, S. M., Sandi, C., Constam, D. & Capsoni, S. 2014. Pathogen-free husbandry conditions alleviate behavioral deficits and neurodegeneration in AD10 anti-NGF mice. *J Alzheimers Dis,* 38**,** 951-64.

Valdes-Mora, F., Song, J. Z., Statham, A. L., Strbenac, D., Robinson, M. D., Nair, S. S., Patterson, K. I., Tremethick, D. J., Stirzaker, C. & Clark, S. J. 2012. Acetylation of H2A.Z is a key epigenetic modification associated with gene deregulation and epigenetic remodeling in cancer. *Genome Res,* 22**,** 307-21.

Valhmu, W. B. 1998. Regulatory Activities of the 5'- and 3'-Untranslated Regions and Promoter of the Human Aggrecan Gene. *Journal of Biological Chemistry,* 273**,** 6196-6202.

Valhmu, W. B., Palmer, G. D., Rivers, P. A., Ebara, S., Cheng, J. F., Fischer, S. & Ratcliffe, A. 1995. Structure of the human aggrecan gene: exon-intron organization and association with the protein domains. *Biochem J,* 309 (Pt 2)**,** 535-42.

Valouev, A., Johnson, S. M., Boyd, S. D., Smith, C. L., Fire, A. Z. & Sidow, A. 2011. Determinants of nucleosome organization in primary human cells. *Nature,* 474**,** 516-20.

Varma, G., Rawat, P., Jalan, M., Vinayak, M. & Srivastava, M. 2015. Influence of a CTCF-Dependent Insulator on Multiple Aspects of Enhancer-Mediated Chromatin Organization. *Mol Cell Biol,* 35**,** 3504-16.

Verfaillie, A., Svetlichnyy, D., Imrichova, H., Davie, K., Fiers, M., Kalender Atak, Z., Hulselmans, G., Christiaens, V. & Aerts, S. 2016. Multiplex enhancer-reporter assays uncover unsophisticated TP53 enhancer logic. *Genome Res*.

Vernimmen, D., Gobbi, M. D., Sloane-Stanley, J. A., Wood, W. G. & Higgs, D. R. 2007. Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. *EMBO J,* 26**,** 2041-51.

Vertel, B. M., Grier, B. L., Li, H. & Schwartz, N. B. 1994. The chondrodystrophy, nanomelia: biosynthesis and processing of the defective aggrecan precursor. *Biochemical Journal,* 301**,** 211-216.

Vertel, B. M., Walters, L. M., Grier, B., Maine, N. & Goetinck, P. F. 1993. Nanomelic chondrocytes synthesize, but fail to translocate, a truncated aggrecan precursor. *Journal of Cell Science,* 104**,** 939-948.

Viapiano, M. S. & Matthews, R. T. 2006. From barriers to bridges: chondroitin sulfate proteoglycans in neuropathology. *Trends Mol Med,* 12**,** 488-96.

Vilim, V. & Fosang, A. J. 1993. Characterization of proteoglycans isolated from associative extracts of human articular cartilage. *Biochem J,* 293 (Pt 1)**,** 165-72.

Vissers, L. E., Lausch, E., Unger, S., Campos-Xavier, A. B., Gilissen, C., Rossi, A., Del Rosario, M., Venselaar, H., Knoll, U., Nampoothiri, S., Nair, M., Spranger, J., Brunner, H. G., Bonafe, L., Veltman, J. A., Zabel, B. & Superti-Furga, A. 2011. Chondrodysplasia and abnormal joint development associated with mutations in IMPAD1, encoding the Golgi-resident nucleotide phosphatase, gPAPP. *Am J Hum Genet,* 88**,** 608-15.

Vorhagen, S., Jackow, J., Mohor, S. G., Tanghe, G., Tanrikulu, L., Skazik-Vogt, C. & Tellkamp, F. 2015. Lineage tracing mediated by cre-recombinase activity. *J Invest Dermatol,* 135**,** e28.

Voros, G., Sandy, J. D., Collen, D. & Lijnen, H. R. 2006. Expression of aggrecan(ases) during murine preadipocyte differentiation and adipose tissue development. *Biochim Biophys Acta,* 1760**,** 1837-44.

Voss, T. C. & Hager, G. L. 2014. Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nat Rev Genet,* 15**,** 69-81.

Walcz, E., Deak, F., Erhardt, P., Coulter, S. N., Fulop, C., Horvath, P., Doege, K. J. & Glant, T. T. 1994. Complete coding sequence, deduced primary structure, chromosomal localization, and structural analysis of murine aggrecan. *Genomics,* 22**,** 364-71.

Wallin, J., Wilting, J., Koseki, H., Fritsch, R., Christ, B. & Balling, R. 1994. The role of Pax-1 in axial skeleton development. *Development,* 120**,** 1109-21.

Wang, J., Gardner, B. M., Lu, Q., Rodova, M., Woodbury, B. G., Yost, J. G., Roby, K. F., Pinson, D. M., Tawfik, O. & Anderson, H. C. 2009. Transcription factor Nfat1 deficiency causes osteoarthritis through dysfunction of adult articular chondrocytes. *J Pathol,* 219**,** 163-72.

Wang, Y., Ma, M., Xiao, X. & Wang, Z. 2012. Intronic splicing enhancers, cognate splicing factors and context-dependent regulation rules. *Nat Struct Mol Biol,* 19**,** 1044-1052.

Watanabe, H., Cheung, S. C., Itano, N., Kimata, K. & Yamada, Y. 1997. Identification of Hyaluronan-binding Domains of Aggrecan. *Journal of Biological Chemistry,* 272**,** 28057-28065.

Watanabe, H., Gao, L., Sugiyama, S., Doege, K., Kimata, K. & Yamada, Y. 1995. Mouse aggrecan, a large cartilage proteoglycan: protein sequence, gene structure and promoter sequence. *Biochem J,* 308 (Pt 2)**,** 433-40.

Watanabe, H., Kimata, K., Line, S., Strong, D., Gao, L. Y., Kozak, C. A. & Yamada, Y. 1994. Mouse cartilage matrix deficiency (cmd) caused by a 7 bp deletion in the aggrecan gene. *Nat Genet,* 7**,** 154-7.

Watanabe, H., Yamada, Y. & Kimata, K. 1998. Roles of aggrecan, a large chondroitin sulfate proteoglycan, in cartilage structure and function. *J Biochem,* 124**,** 687-93.

Watanabe, Y., Takeuchi, K., Higa Onaga, S., Sato, M., Tsujita, M., Abe, M., Natsume, R., Li, M., Furuichi, T., Saeki, M., Izumikawa, T., Hasegawa, A., Yokoyama, M., Ikegawa, S., Sakimura, K., Amizuka, N., Kitagawa, H. & Igarashi, M. 2010. Chondroitin sulfate N-acetylgalactosaminyltransferase-1 is required for normal cartilage development. *Biochem J,* 432**,** 47-55.

Weedon, M. N., Lango, H., Lindgren, C. M., Wallace, C., Evans, D. M., Mangino, M., Freathy, R. M., Perry, J. R., Stevens, S., Hall, A. S., Samani, N. J., Shields, B., Prokopenko, I., Farrall, M., Dominiczak, A., Johnson, T., Bergmann, S., Beckmann, J. S., Vollenweider, P., Waterworth, D. M., Mooser, V., Palmer, C. N., Morris, A. D., Ouwehand, W. H., Zhao, J. H., Li, S., Loos, R. J., Barroso, I., Deloukas, P., Sandhu, M. S., Wheeler, E., Soranzo, N., Inouye, M., Wareham, N. J., Caulfield, M., Munroe, P. B., Hattersley, A. T., Mccarthy, M. I. & Frayling, T. M. 2008. Genome-wide association analysis identifies 20 loci that influence adult height. *Nat Genet,* 40**,** 575-83.

Weiss, D., Liggitt, D. & Clark, J. 1999. Histochemical Discrimination of Endogenous Mammalian β-galactosidase Activity from that Resulting from lac-Z Gene Expression. *The Histochemical Journal,* 31**,** 231-236.

Wenger, A. M., Clarke, S. L., Notwell, J. H., Chung, T., Tuteja, G., Guturu, H., Schaar, B. T. & Bejerano, G. 2013. The enhancer landscape during early neocortical development reveals patterns of dense regulation and co-option. *PLoS Genet,* 9**,** e1003728.

Werner, T., Hammer, A., Wahlbuhl, M., Bosl, M. R. & Wegner, M. 2007. Multiple conserved regulatory elements with overlapping functions determine Sox10 expression in mouse embryogenesis. *Nucleic Acids Res,* 35**,** 6526-38.

Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., Rahl, P. B., Lee, T. I. & Young, R. A. 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell,* 153**,** 307-19.

Wilson, R., Diseberg, A. F., Gordon, L., Zivkovic, S., Tatarczuch, L., Mackie, E. J., Gorman, J. J. & Bateman, J. F. 2010. Comprehensive profiling of cartilage extracellular matrix formation and maturation using sequential extraction and label-free quantitative proteomics. *Mol Cell Proteomics,* 9**,** 1296-313.

Wu, Y., Liang, D., Wang, Y., Bai, M., Tang, W., Bao, S., Yan, Z., Li, D. & Li, J. 2013. Correction of a genetic disease in mouse via use of CRISPR-Cas9. *Cell Stem Cell,* 13**,** 659-62.

Wu, Y., Zhou, H., Fan, X., Zhang, Y., Zhang, M., Wang, Y., Xie, Z., Bai, M., Yin, Q., Liang, D., Tang, W., Liao, J., Zhou, C., Liu, W., Zhu, P., Guo, H., Pan, H., Wu, C., Shi, H., Wu, L., Tang, F. & Li, J. 2015. Correction of a genetic disease by CRISPR-Cas9-mediated gene editing in mouse spermatogonial stem cells. *Cell Res,* 25**,** 67-79.

Wuidart, A., Ousset, M., Rulands, S., Simons, B. D., Van Keymeulen, A. & Blanpain, C. 2016. Quantitative lineage tracing strategies to resolve multipotency in tissue-specific stem cells. *Genes Dev,* 30**,** 1261-77.

Yamagata, M., Shinomura, T. & Kimata, K. 1993. Tissue variation of two large chondroitin sulfate proteoglycans (PG-M/versican and PG-H/aggrecan) in chick embryos. *Anat Embryol (Berl),* 187**,** 433-44.

Yamaguchi, Y. 2000. Lecticans: organizers of the brain extracellular matrix. *Cell Mol Life Sci,* 57**,** 276-89.

Yan, J., Zhang, L., Xu, J., Sultana, N., Hu, J., Cai, X., Li, J., Xu, P. X. & Cai, C. L. 2014. Smad4 Regulates Ureteral Smooth Muscle Cell Differentiation during Mouse Embryogenesis. *PLoS One,* 9**,** e104503.

Yang, H., Wang, H., Shivalila, C. S., Cheng, A. W., Shi, L. & Jaenisch, R. 2013. One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell,* 154**,** 1370-9.

Yang, L., Tsang, K. Y., Tang, H. C., Chan, D. & Cheah, K. S. 2014. Hypertrophic chondrocytes can become osteoblasts and osteocytes in endochondral bone formation. *Proc Natl Acad Sci U S A,* 111**,** 12097-102.

Yang, R., Kerschner, J. L., Gosalia, N., Neems, D., Gorsic, L. K., Safi, A., Crawford, G. E., Kosak, S. T., Leir, S. H. & Harris, A. 2016. Differential contribution of cis-regulatory elements to higher order chromatin structure and expression of the CFTR locus. *Nucleic Acids Res,* 44**,** 3082-94.

Yao, L., Berman, B. P. & Farnham, P. J. 2015. Demystifying the secret mission of enhancers: linking distal regulatory elements to target genes. *Critical Reviews in Biochemistry and Molecular Biology,* 50**,** 550-573.

Yeung Tsang, K., Wa Tsang, S., Chan, D. & Cheah, K. S. 2014. The chondrocytic journey in endochondral bone growth and skeletal dysplasia. *Birth Defects Res C Embryo Today,* 102**,** 52-73.

Yin, H., Xue, W., Chen, S., Bogorad, R. L., Benedetti, E., Grompe, M., Koteliansky, V., Sharp, P. A., Jacks, T. & Anderson, D. G. 2014. Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nat Biotechnol,* 32**,** 551-3.

Yoshida, M., Hata, K., Takashima, R., Ono, K., Nakamura, E., Takahata, Y., Murakami, T., Iseki, S., Takano-Yamamoto, T., Nishimura, R. & Yoneda, T. 2015. The transcription factor Foxc1 is necessary for Ihh-Gli2-regulated endochondral ossification. *Nat Commun,* 6**,** 6653.

Yu, J. & Urban, J. P. G. 2010. The elastic network of articular cartilage: an immunohistochemical study of elastin fibres and microfibrils. *Journal of Anatomy,* 216**,** 533-541.

Yu, L., Liu, H., Yan, M., Yang, J., Long, F., Muneoka, K. & Chen, Y. 2007. Shox2 is required for chondrocyte proliferation and maturation in proximal limb skeleton. *Dev Biol,* 306**,** 549-59.

Yusufzai, T. M., Tagami, H., Nakatani, Y. & Felsenfeld, G. 2004. CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Mol Cell,* 13**,** 291-8.

Zanin, M. K., Bundy, J., Ernst, H., Wessels, A., Conway, S. J. & Hoffman, S. 1999. Distinct spatial and temporal distributions of aggrecan and versican in the embryonic chick heart. *Anat Rec,* 256**,** 366-80.

Zanotti, S. & Canalis, E. 2013. Notch suppresses nuclear factor of activated T cells (NFAT) transactivation and Nfatc1 expression in chondrocytes. *Endocrinology,* 154**,** 762-72.

Zeng, L., Kempf, H., Murtaugh, L. C., Sato, M. E. & Lassar, A. B. 2002. Shh establishes an Nkx3.2/Sox9 autoregulatory loop that is maintained by BMP signals to induce somitic chondrogenesis. *Genes Dev,* 16**,** 1990-2005.

Zhang, C., Yang, F., Cornelia, R., Tang, W., Swisher, S. & Kim, H. 2011. Hypoxia-inducible factor-1 is a positive regulator of Sox9 activity in femoral head osteonecrosis. *Bone,* 48**,** 507-13.

Zhang, Y., Rajan, R., Seifert, H. S., Mondragon, A. & Sontheimer, E. J. 2015. DNase H Activity of Neisseria meningitidis Cas9. *Mol Cell,* 60**,** 242-55.

Zhao, C., Li, X. & Hu, H. 2016. PETModule: a motif module based approach for enhancer target gene prediction. *Sci Rep,* 6**,** 30043.

Zhao, H. & Dean, A. 2004. An insulator blocks spreading of histone acetylation and interferes with RNA polymerase II transfer between an enhancer and gene. *Nucleic Acids Res,* 32**,** 4903-19.

Zhao, J., Herrera-Diaz, J. & Gross, D. S. 2005. Domain-wide displacement of histones by activated heat shock factor occurs independently of Swi/Snf and is not correlated with RNA polymerase II density. *Mol Cell Biol,* 25**,** 8985-99.

Zheng, Q., Keller, B., Zhou, G., Napierala, D., Chen, Y., Zabel, B., Parker, A. E. & Lee, B. 2009. Localization of the cis-enhancer element for mouse type X collagen expression in hypertrophic chondrocytes in vivo. *J Bone Miner Res,* 24**,** 1022-32.

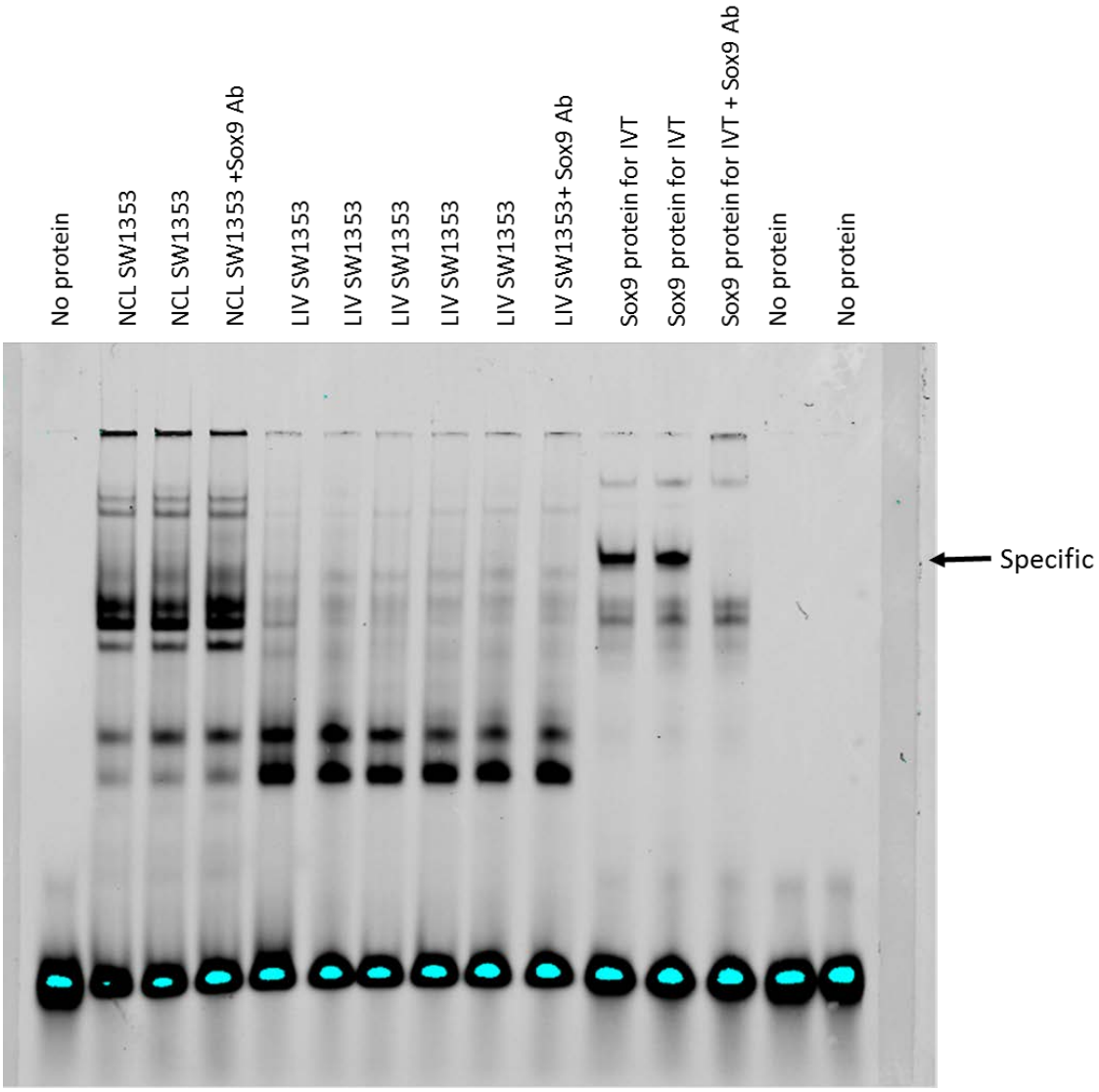
Zhou, G., Lefebvre, V., Zhang, Z., Eberspaecher, H. & De Crombrugghe, B. 1998. Three high mobility group-like sequences within a 48-base pair enhancer of the Col2a1 gene are required for cartilage-specific expression in vivo. *J Biol Chem,* 273**,** 14989-97.

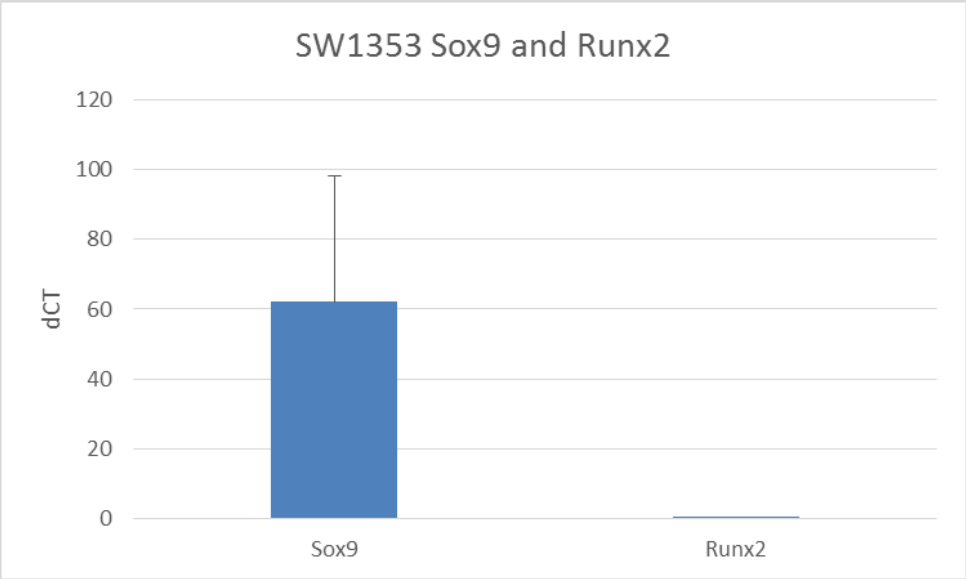
Zhou, G., Zheng, Q., Engin, F., Munivez, E., Chen, Y., Sebald, E., Krakow, D. & Lee, B. 2006. Dominance of SOX9 function over RUNX2 during skeletogenesis. *Proc Natl Acad Sci U S A,* 103**,** 19004-9.

Zhou, X., Von Der Mark, K., Henry, S., Norton, W., Adams, H. & De Crombrugghe, B. 2014. Chondrocytes transdifferentiate into osteoblasts in endochondral bone during development, postnatal growth and fracture healing in mice. *PLoS Genet,* 10**,** e1004820.

7. APPENDICES

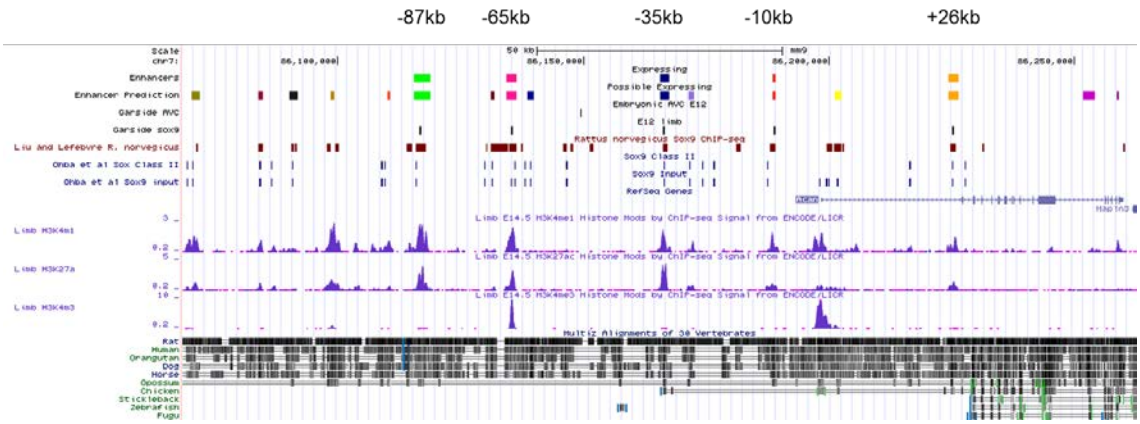
Figure 84: Direct EMSA comparing nuclear protein extracts from SW1353 cells and qPCR for Sox9 in SW1353





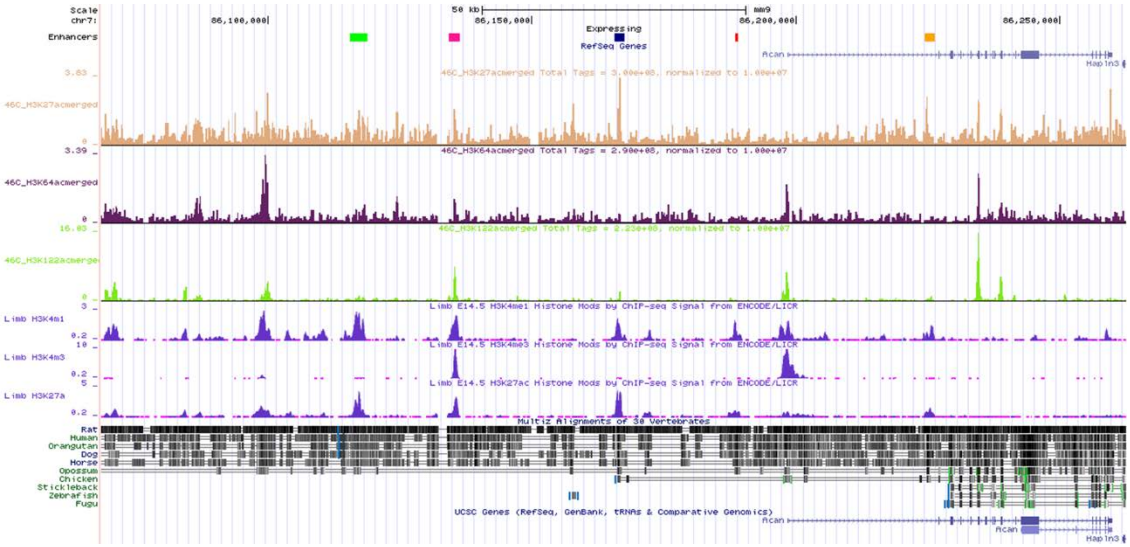
EMSA using nuclear protein extracted from SW1353 from Liverpool and Newcastle are ran alongside the in vitro transcribed Sox9 protein. Super shifts were performed with Sox9 anti-body and reveals that the size of Sox9 is different in cells from different areas, in fact there is no supershifts in any of the SW1353 lysates. The shift was generated by Dr L. Reynard at Newcastle University using probes and lysates generated at University of Liverpool. On an mRNA level there is actually Sox9 present in SW1353 compared to the Runx2.

Figure 85: UCSC genome browser output with custom tracks from Sox9 ChIP-Seq



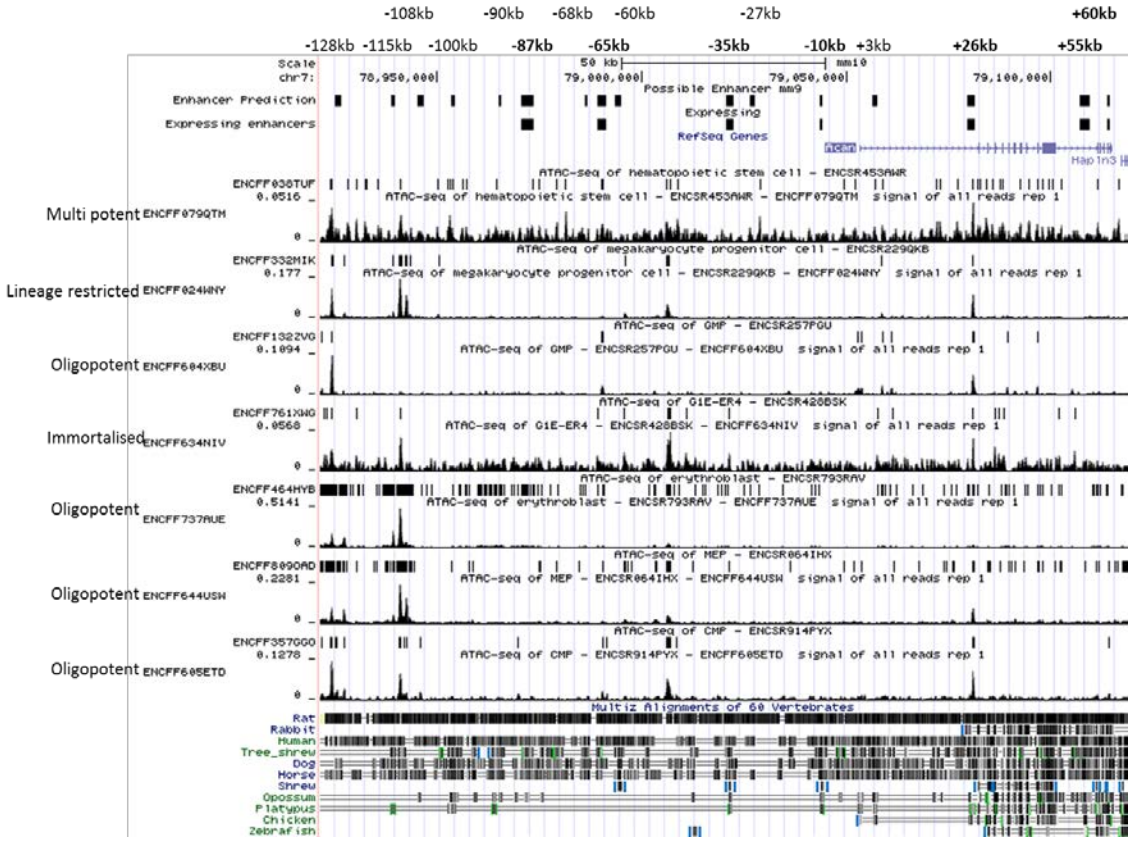
Sox9 ChIP-seq from different groups show that Acan interacts with Sox9 frequently. The identified Acan enhancers that express in the chondrocyte are the only sites to interact with Sox9 in embryonic mouse limbs at E12. In mouse chondrocytes the Acan enhancers interact in class I and II interactions and bind in rat chondrocytes.

Figure 86: Analysis of histone modifications in the *Acan* locus that mark different types of enhancer



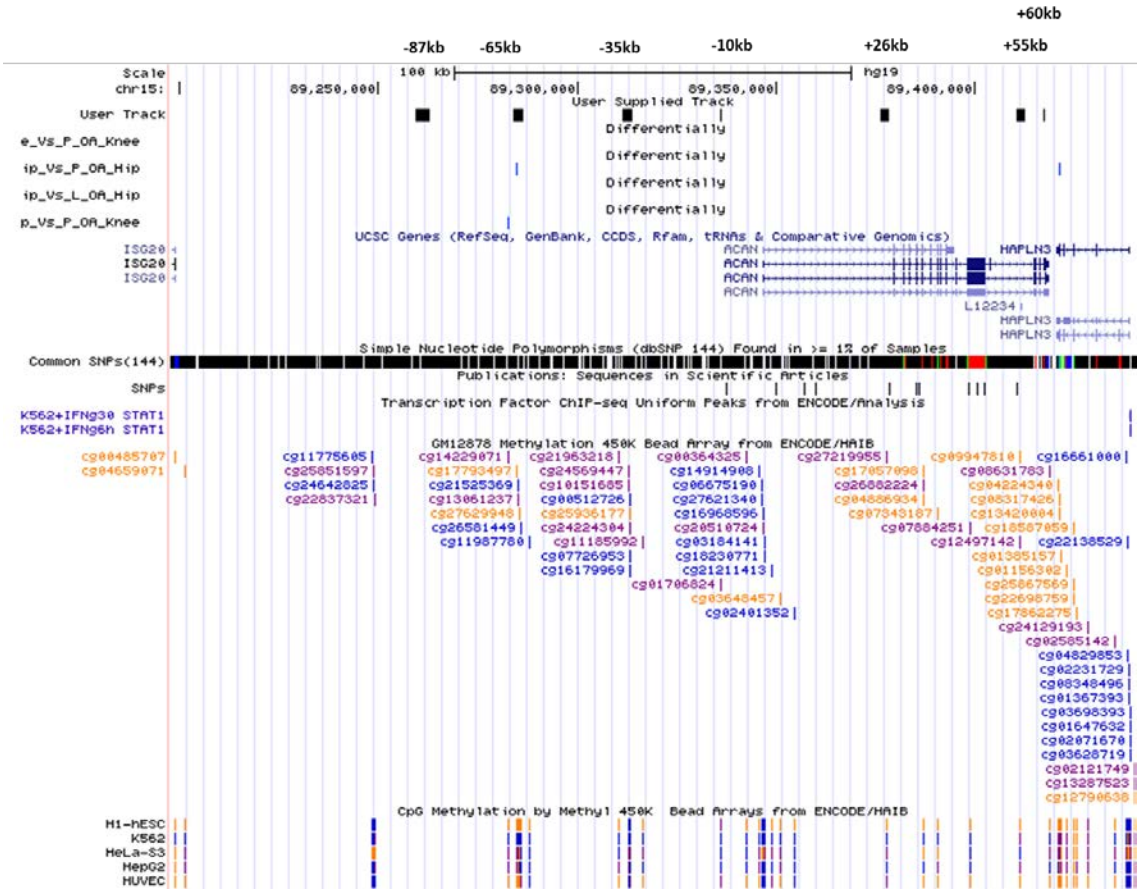
New enhancer markers are being identified to aid in the identification of enhancer elements. Active enhancers different to those marked by H3K27ac have been characterised. The new class of active enhancers are marked by H3K64ac of which the Acan enhancers identified in this study do not overlap with the exception from the -65kb region which has a minor peak.

Figure 87: ATAC-seq of the *Acan* locus in haematopoietic cells



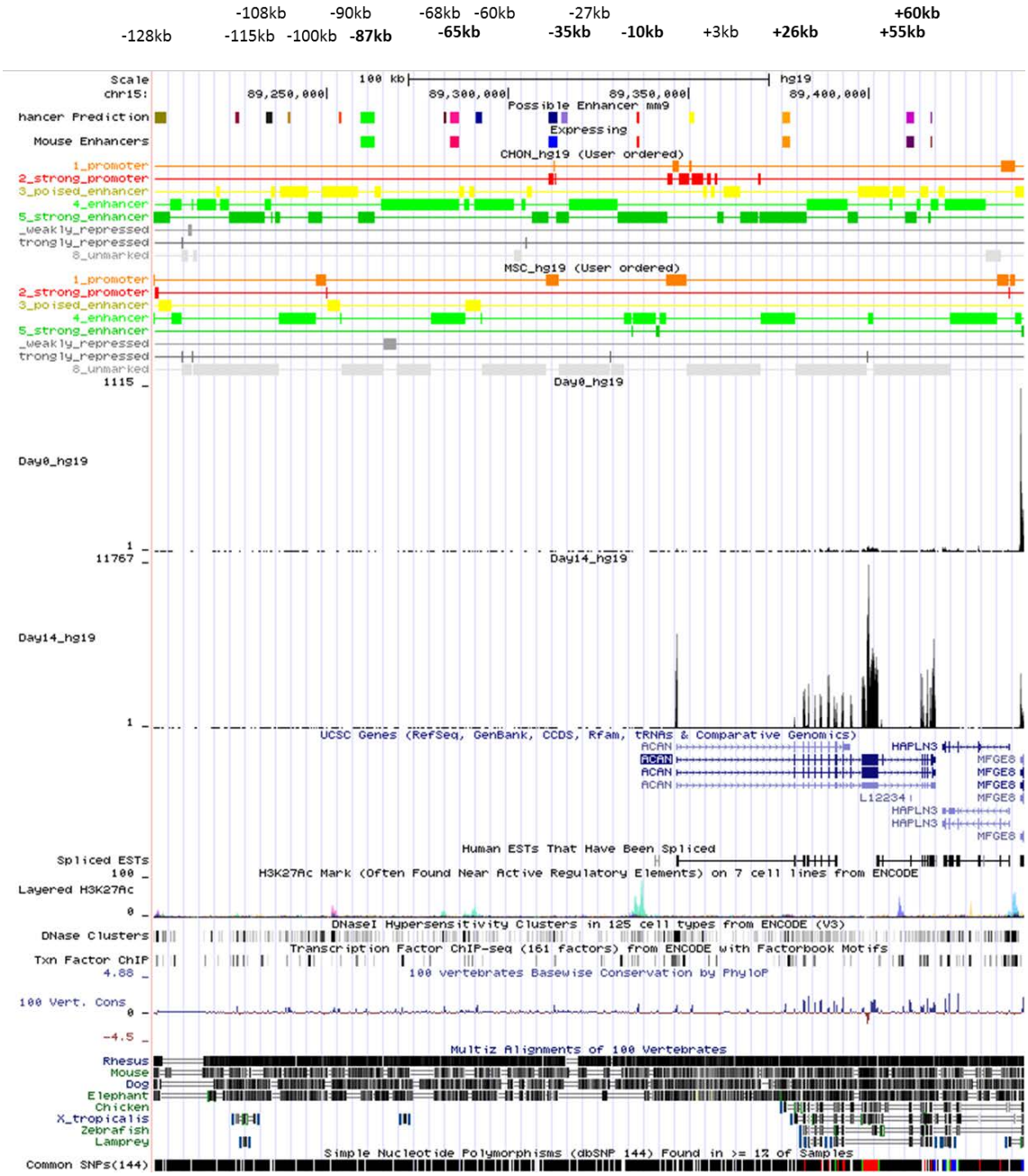
ATAC-seq from haematopoietic cells allows an assessment of the chromatin landscape of Acan. Areas of open chromatin are marked by peaks. Only the +26kb appears to be open in these cell types.

Figure 88: Methylation data in hip and knee OA for the *Acan* locus



DNA methylation differentially expressed regions. The -65kbis differentially methylated in OA hips compared to normal hips. Other regions do not appear to be affected by methylation in human OA. Data provided by Dr. L. Reynard.

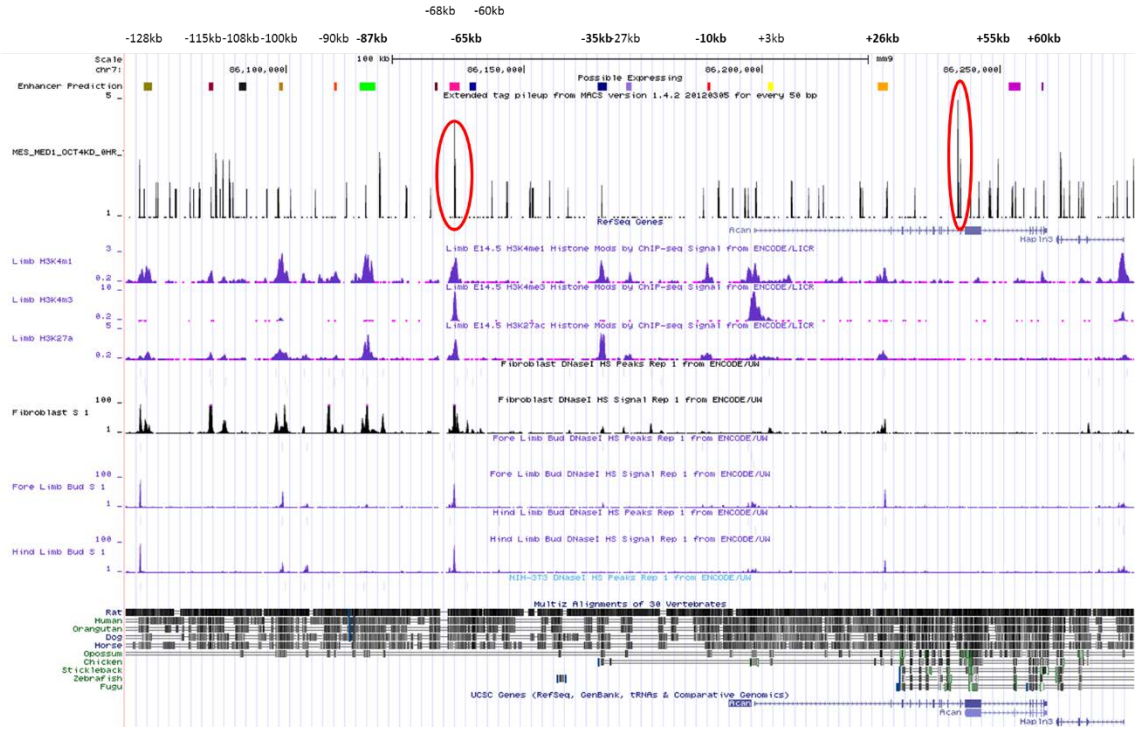
Figure 89: Histone changes when MSC differentiate to chondrocytes in the *Acan* locus



Gene expression and collated ChIP-data for genome state from MSC and 14 days of differentiation. Acan is only transcribed in chondrocytes. Each one of the Acan enhancers identified in this study shows enhancer activity in MSC with the exception of the -87kb which is inactive and the -35kb which are marked as a promoter. In chondrocytes the

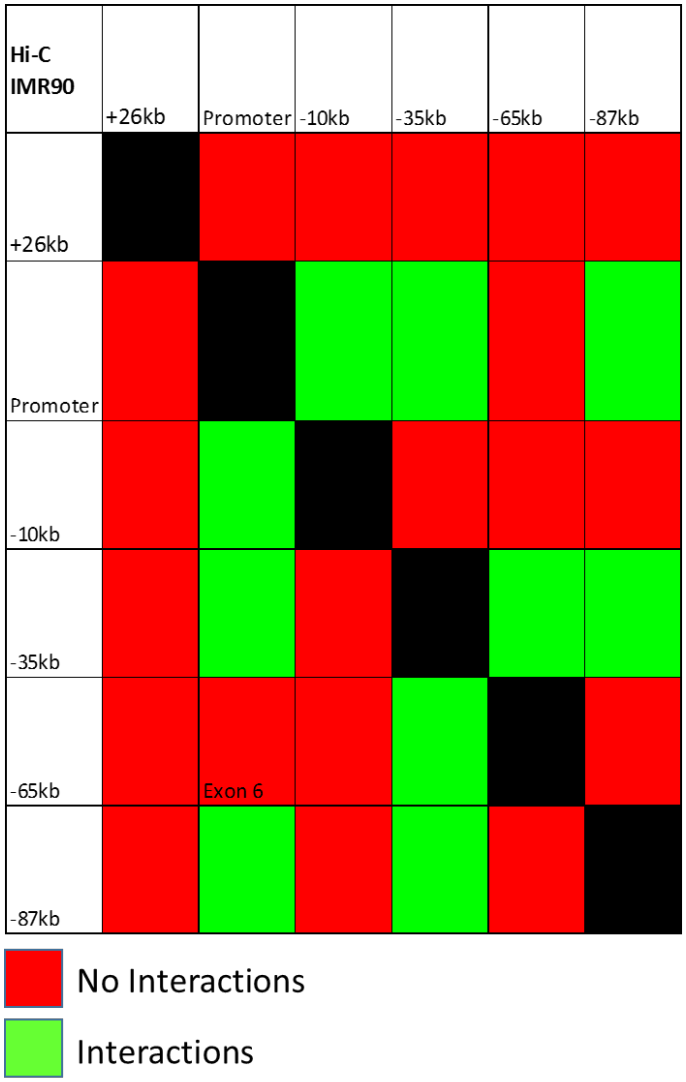
+26kb and -10kb become strong enhancers, the -65kb remains an enhancer, the -35kb becomes a strong promoter and the -87kb is only active in chondrocytes as a strong enhancers suggesting that they play a role in the chondrocyte lineage determination or the tissue specificity of the Acan. Data provided by K. Cheung and Prof. D. Young.

Figure 90: Med1 interaction sites in the *Acan* locus



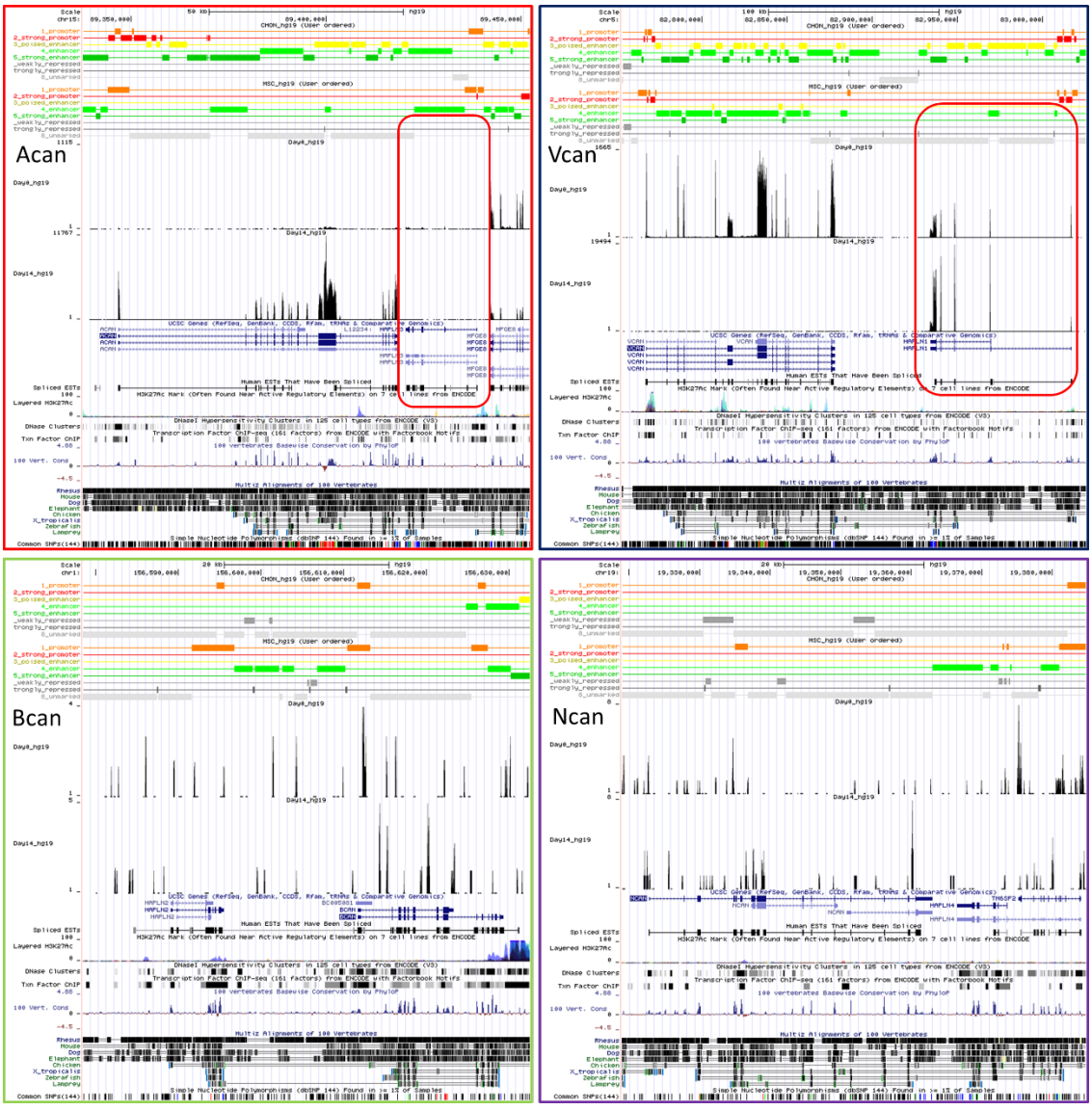
Data from Med1 ChIP-seq. Med1 is a subunit of the mediator that provides a bridge between DNA elements and the promoter to activate transcription. The Med1 localises strongly with the -65kb enhancer element and the 9th exon of Acan (red). Suggesting that the -65kb plays an important role in the activation of Acan transcription.

Figure 91: Interactions of the *Acan* enhancers in IMR-90 human lung cells from 4D genomes



4D genomes output for the Acan region based on Hi-C. Interactions with other enhancers by the +26kb do not occur in the IMR-90 but the regions does interact with HAPLN3. The promoter (chr15:89,346,048-89,347,590) interacts with the -10kb, -35kb and -87kb. The -10kb only interacts with the promoter, the -35kb interacts with the promoter, -65kb and -87kb. The -65kb interacts with Exon 6 of Acan and the -35kb. The -87kb interacts with the promoter and the -35kb. We cannot discount other ling range interaction, data generated from: http://4dgenome.research.chop.edu/ .

Figure 92: RNA expression of MSC differentiated into chondrocytes for different Lectican proteins



Gene profiles for transcription in the lectican and link protein families in the chondrocytes and MSCs. Acan is active in the chondrocyte but HAPLN3 does not, Vcan is active in the chondrocyte but HAPLN1 expression is reduced in the chondrocyte. Bcan and Ncan are neuronal specific and displays an erratic expression in the chondrocytes and this is reflected win HAPLN2 and HAPLN4.

"Not until they set aside their differences did I see the true power they all share deep inside. I see now that the circumstances of one's birth are irrelevant; it is what you do with the gift of life that determines who you are." – M. 1998.